(12)

NUWC-NPT Technical Report 10,222
1 September 1993

Propagation of Evidence Through Fuzzy Rules

P. R. Kersten
Combat Control Systems Department

Naval Undersea Warfare Center Division
Newport, Rhode Island

PREFACE

This report was funded under NUWC Newport IR/IED Project No. 802424 "Fuzzy Expert Systems." The IR/IED program is funded by the Office of Naval Research; the NUWC Newport program manager is K. M. Lima (Code 102).

The technical reviewer for this report was K. F. Gong (Code 2211).

The author gratefully acknowledges the enthusiastic support of K. F. Gong, S. E. Hammel, and W. G. Ravo for the Fuzzy Systems effort in Code 2211. The author is thankful to Professor Avi Kak of Purdue University for some helpful comments on this report.

Reviewed and Approved: 1 September 1993

P. A. La Brecque

P. A. La Brecque
Head, Combat Control Systems Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
1 September 1993

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Propagation of Evidence Through Fuzzy Rules

5. FUNDING NUMBERS

6. AUTHOR(S)
P. R. Kersten

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Undersea Warfare Center Division
1176 Howell St.
Newport, RI 02841-1708

8. PERFORMING ORGANIZATION REPORT NUMBER
TR 10,222

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Propagation of evidence through fuzzy rules is discussed and illustrated for three different certainty representations. The first certainty representation is single-valued using a real number to represent the certainty for both the premise and the conclusion. This single-valued representation is often found in fuzzy expert system shells. The second certainty representation is interval-valued and is useful in expert system applications. Finally, the third certainty representation is functional-valued and is the most general and the most complex.

14. SUBJECT TERMS
Fuzzy Logic
Fuzzy Rules
Expert Systems
Certainty

15. NUMBER OF PAGES
42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

FES	Fuzzy Expert System
FLOPS	Fuzzy Logic Official Production System
MPG	*Modus ponens* generating
TV	Trillas and Valverde
T-norm	Triangular norm
S-norm	Triangular conorm
CRI	Compositional Rule of Inference
CDF	Cumulative distribution function
PDF	Probability density function

PROPAGATION OF EVIDENCE THROUGH FUZZY RULES

INTRODUCTION

Most fuzzy system models are based on the principle of embedding (reference 1), which takes the problem to be solved, embeds it in a richer representation space, solves the problem in this new space, and then projects the solution back into the desired output space. This principle is a powerful technique often employed in mathematical analysis. First, consider a classical example of this technique and then an example of the same technique applied to fuzzy systems to illustrate not only how fuzzy rules are used to solve system problems, but also how evidence is propagated through the fuzzy rules. Evidence means the degree of certainty that the data satisfy the premise of the fuzzy rule.

One example of this classical embedding technique is the integration of improper real integrals (reference 2). First the integrand is complexified (i.e., the variable of integration is replaced by a complex variable) so that it can now be embedded in the complex number domain; complex numbers are a far richer representation than real variables. Then a closed path of integration is chosen so that the real number line is included in the path. The residue theorem is then employed to evaluate the integral about the closed path, and the line integral along one part of the integration path is then the value desired. Decomplexification is trivial since it amounts to summing all the other components of the integral path to obtain the desired solution. Note in this case, not only is the problem embedded in a far richer field of numbers, it is also embedded in a far more complex integration path. Other examples of this can be found in Bezdek's paper (reference 1).

In fuzzy models, a problem is embedded in a fuzzy rule base system by first fuzzifying the input, solving the problem using fuzzy rules, and defuzzifying to project back into the solution space. Figure 1 illustrates the overall block diagram of this system. Fuzzifying the data amounts to mapping the input variables into linguistic variables, which are defined in terms of fuzzy sets; the linguistic variable called COLOR is an example and is illustrated in figure 2. Radiation of a given frequency is translated into linguistic terms with a membership value. The frequency marked in this diagram has a 0.3 membership in both GREEN and BLUE. The fuzzy sets defined on the base variable of frequency define the term set, which in this case consists of {RED, ORANGE, YELLOW, GREEN, BLUE, INDIGO, VIOLET}.

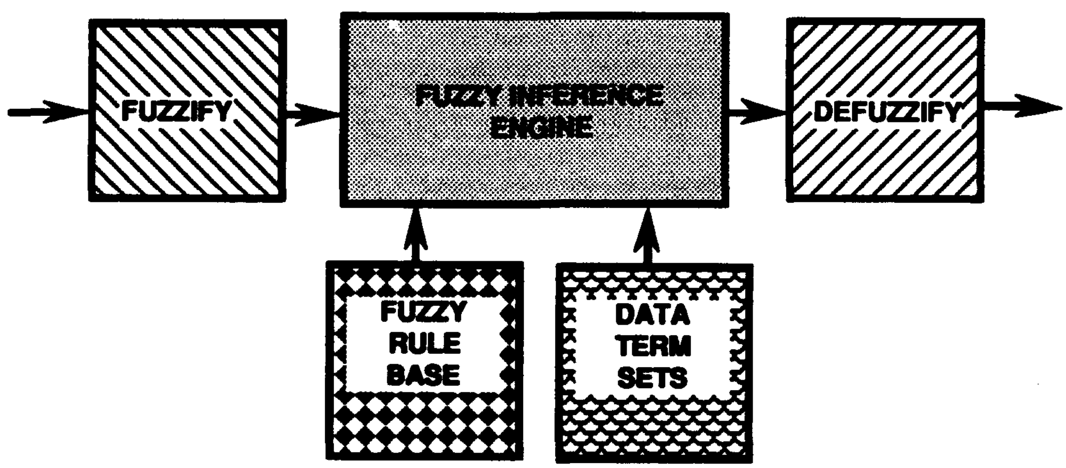


Figure 1. Fuzzy System Model

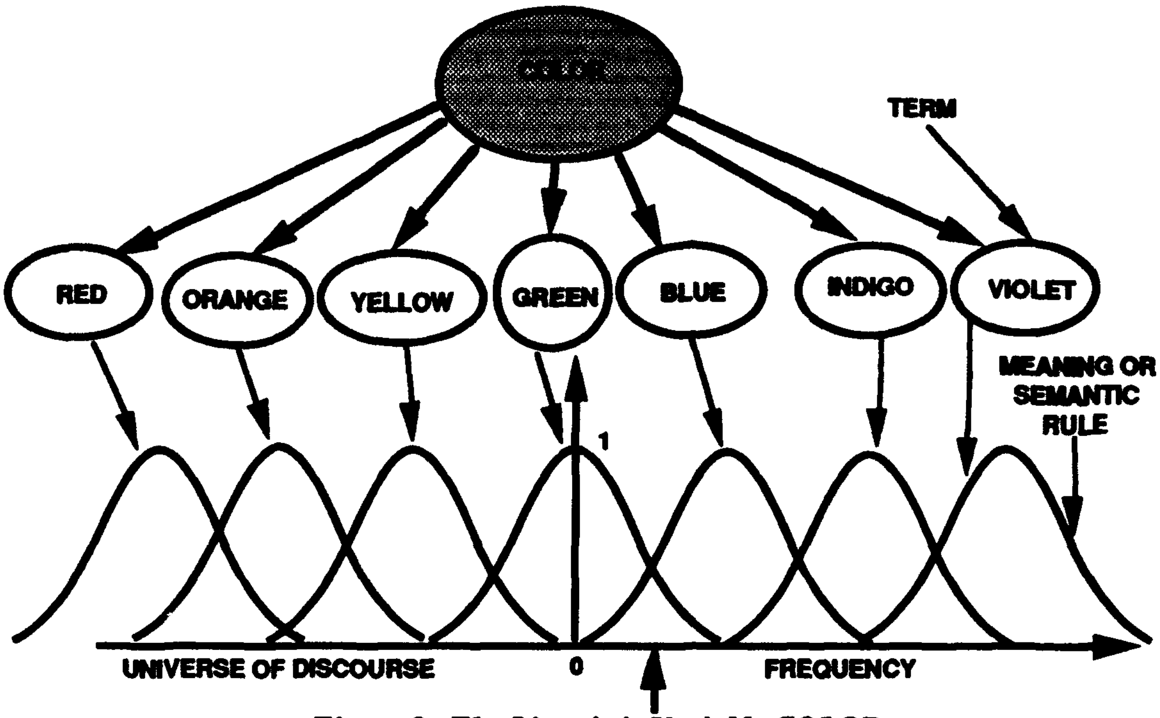


Figure 2. The Linguistic Variable COLOR

The linguistic variables used in the typical fuzzy control system are usually sensor values. Figure 3 illustrates how a control system called a taxi driver might determine the rate of braking in approaching a red light. Note here that the rate of braking is dependent on the speed of the automobile and the relative distance to the stopped cars. Physical laws of momentum dictate that the faster the vehicle is going and the closer the stop light, then the harder one should brake. For the specific case considered in figure 3, the entire model consists of two inputs and two rules. The fuzzy rules have the following form:

IF the speed of the car is high,
AND the distance to the stop light is near,
THEN the braking should be hard.

IF the speed of the car is medium,
AND the distance to the stop light is medium,
THEN the braking should be medium.

Note that the figures are sketches of the solutions, not exact calculated values. The two braking rules yield different conclusions, which are aggregated to yield a single output braking rate. This aggregation procedure is called defuzzification and is a simple averaging of the areas in the output fuzzy sets. Rule conclusion strengths are set equal to the minimum degree of membership that the inputs have in the premise clauses. In effect, the strength of the output for a rule is determined by the degree of satisfaction of the premise clauses. Premise satisfaction or certainty and its propagation through the rules are an inherent part of the solution technique. In fact, the input data are evidence only if the data are relevant to the rules, and relevance is equated to data satisfaction of the premise measured by the certainty of the premise. Propagation of the certainty through the ply to determine the certainty of the conclusion is the propagation of evidence through the fuzzy rule.

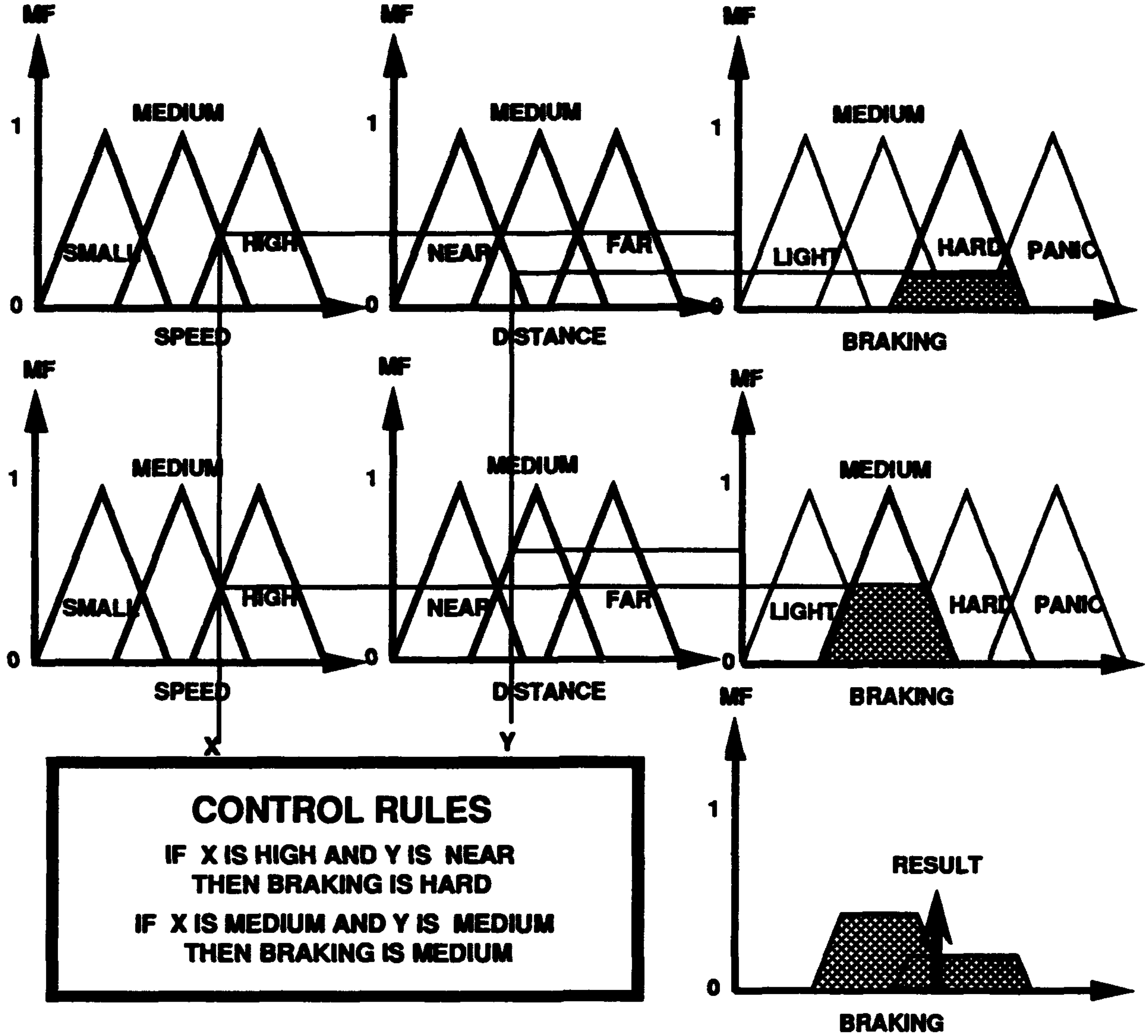


Figure 3. Typical Output Strength Calculation in Fuzzy Control Logic

Figure 3 contains the same basic components shown in the general system of figure 1. Fuzzification takes place when the sensor values are used as inputs to the term sets or fuzzy membership functions. The rule base consists of two rules in this small example, and the term sets are all drawn out using triangular term sets. The inference engine combines the premise certainties to find the certainty of the conclusion, which is represented by truncating the conclusion membership function. Defuzzification is the method of aggregating the conclusions to find the resulting control value.

In the above example, the certainty of the premise or its validity or, equivalently, its degree of satisfaction is a single-valued real number. Here, the notion of strength is equivalent to the notion of certainty value or validity. This process is only one way to determine the certainty of the premise clauses, the premise, and finally the certainty of the output. Certainty can be represented

as a single-valued real number, a certainty interval, or a linguistic variable. When the fuzzy model is applied to classification problems, the decisions are more open loop, i.e., their validity is not immediately tested by the system. In effect, there is a larger delay in the feedback loop so it is more important to know the certainty of the decisions. Here, conclusion certainty is more important and thus more complex representations better describe the conclusion validity.

The braking example uses a standard defuzzification rule or conclusion aggregation technique. Other applications require more general aggregation techniques. This example does not model the strength of the rule itself. Fuzzy rules can associate certainty with the ply. Here, certainty represents the designer's faith in the rule. This certainty will be factored in the propagation schemes considered in the next sections.

As outlined above, three certainty representation schemes are considered in this report. Ordered by representational complexity they are as follows: (1) A single-valued measure of certainty, (2) an interval-valued measure of certainty, and (3) a functional-valued measure of certainty. Several fuzzy expert system (FES) shells use the single-valued certainty measure. Part of the popularity of this measure is its simplicity and practicality. A single-valued certainty is associated with the premise, with the ply, and with the consequence. A simple functional combination of the premise and implication certainty produces the consequence certainty. Hall and Kandel (reference 3) use a single-valued evaluation of the premise and the ply, but the functional propagation to the consequence is no longer simple, since it depends on the functional form of the ply. These results are based on Trillas and Valverde's (reference 4) method of certainty propagation using a single-valued validity measure.

The second scheme uses interval-valued measures of certainty so that both the premise and the conclusion have certainty intervals associated with them. The certainty of the ply is represented by a pair of numbers indicating the strength of the ply in the forward and reverse directions, repectively. The certainty of the conclusion is derived from the certainty of the premise and the certainty of the ply. For each rule, the system designer must supply the certainty of the ply and calculate from the data the certainty interval for the premise. A proponent of this interval-valued method is Piero Bonissone (references 5-7). Bounds on the premise are generated from the data using possibility theory. This interval method also can aggregate the conclusion certainty, which determines the conclusion certainty when several rules reach the same conclusion but with differing intervals of certainty.

The third method is a generalization of the certainty representation that employs the Fuzzy Inclusion Index (or index as it will be used in this report) to determine the truth of a fuzzy predicate. Thus, truth is represented by the membership function of a fuzzy set, which is compared with the terms of the linguistic variable called TRUTH. This approach was introduced first by Zadeh (reference 8). This fuzzy certainty measure requires more sophisticated methods for both propagation and interpretation. The index contains more information than is contained in either the interval-valued or the single-valued certainty representation. Appendix A discusses the Fuzzy Inclusion Index in detail.

In what follows, three different certainty representations are discussed. The single-valued representation as illustrated in the braking example, the interval-valued representation that has the look and feel of a confidence interval, and the linguistic variable representation where the certainty is a fuzzy set. With each representation, the propagation of the certainty through the implication is presented and illustrated. The best choice for the appropriate representation and propagation scheme is a function of the application. Emphasis in this report is placed on classification problems, so some conclusions for this type of problem will be drawn.

SINGLE-VALUED CERTAINTY PROPAGATION

This section considers a single-valued certainty representation and its propagation through the implication operator (the ply). There are many different methods for propagation of evidence through the ply, but only two are discussed here. One method can be found in the expert system shell Fuzzy Logic Official Production System (FLOPS) as reported in Buckley (references 9-12). The second method uses the work of Trillas and Valverde (reference 4). Both methods are discussed below.

Before discussing the propagation of evidence through the implication operator, it behooves us to discuss why one must explicitly represent the certainty of the ply. In the previous section, the fuzzy rules had conclusions whose strength was determined as the minimum of the two premise clauses. The rule was assumed to be absolute so if the premise was satisfied with certainty one, the conclusion has certainty one. However, not all rules are absolute and the certainty associated with the ply itself tries to model this; e.g., suppose the rule states *if you elect me, then I will lower your taxes,* then one must really model the strength of the ply. Even for physical models, the conditions may be so uncertain that normal physical laws must be asserted with reservation. Thus the explicit representation of the implication certainty is a critical component of the propagation of evidence.

The first method, a simple propagation scheme, determines the certainty of the conclusion from the certainty of the premise and the ply as follows:

$$m(v(a), v(a \to b)) = \min(v(a), v(a \to b)).$$

This formula states that the certainty of the conclusion is the minimum of the certainty attributed to the ply and the premise. The function $v(a)$ stands for the validity of the clause called a, and $v(a \to b)$ is the validity of the rule $a \to b$. The validity function is a mapping from the set of clauses to the interval [0,1]. Validity is Trillas and Valverde's terminology, and one can interpret this to mean the degree of truth or the certainty. In this report, certainty and validity will be used interchangeably, but note that the validity is not the same as the degree of membership in a fuzzy set, except in special cases where the input values are known exactly. In the previous section, the sensor values were used as arguments to the term sets, and the resulting membership functions were interpreted as certainties or validities. This special case is important, but if the sensor readings are fuzzy sets, then this interpretation is meaningless. In these cases, measure of the overlap and subsethood of the sensor reading with respect to the term sets must be used to determine the validity of the premise or its clauses.

Premise evaluation uses the minimum of the certainties for conjunctions of clauses forming the premise. Thus, when the premise is a conjunction of clauses of the form

$$a = \bigcap_{i=1}^{n} a_i$$

and $v(a_i)$ is the truth associated with each of these clauses, then $v(a) = \min_{0 \le i \le n} v(a_i)$.
Disjunctions in the premise can be handled as the maximum of the certainties of the clauses making up a premise (reference 9, p. 6). Thus, the confidence in the conclusion is the minimum of the validity of the rule antecedent and ply.

Evaluating the validity of the premise is the subject of a vast amount of literature. The validity of the ply is under the purview of the system designer, so its value is assumed to be known or at least estimated, and the validity of the premise is something that can be calculated from the data and the fuzzy sets used to represent the premise. Another requirement of the designer is to supply the matching algorithm for comparing the data with the premise. In this report, matching is based on the necessity and the possibility, which are defined later. Other matching algorithms are discussed more thoroughly in appendix B. Thus, ply validities are assumed to be known, and the premise validities are calculated from the data and the clauses that make up the premise.

Figure 4 illustrates the single-valued certainty propagation algorithm for three distinctly different fuzzy data inputs. For each case, the certainty value is defined to be the possibility Π, which is $\Pi = \sup_{x \varepsilon X} \min[\mu_A(x), \mu_B(x)] = \sup_{x \varepsilon X} \mu_A(x) \wedge \mu_B(x)$ where $\mu_A(x) \wedge \mu_B(x)$ is defined to be the $\min[\mu_A(x), \mu_B(x)]$. The figure illustrates one problem associated with this approach. A single-valued certainty is not sufficient to account for the overlap and spread of the fuzzy sets representing the data and the premise. The three cases illustrated all yield the same single-valued certainty, the degree of overlap of the fuzzy data set A and the fuzzy fact set B. Yet the degree that the data set is a subset of the fact set is clearly quite different for each of the cases. The single-valued certainty representation can only capture one facet of the match of the data to the premise. The strength of the single-valued certainty representation is its simplicity and also its weakness since one facet of the certainty is not sufficient to describe how the data match the premise. In this example, the single-valued certainty illustrates the overlap but does not capture the subsethood of the data to the fact. With only one value to represent the matching of the data to the premise, it is very important to choose this value carefully, so that it summarizes the important system feature for the particular application.

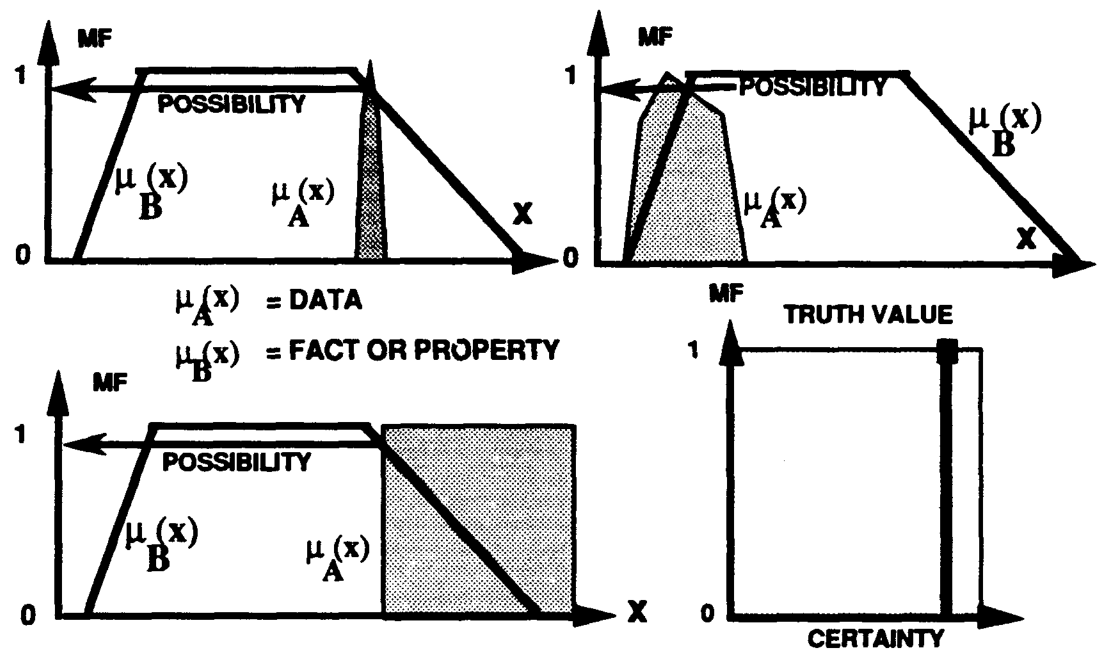


Figure 4. Three Examples that Have the Same Single-Valued Certainty

Note that the structure of the ply is essentially ignored, representing its strength as a real number. In multivalued logic, there are many different plys named after famous logicians. Assigning a certainty to the ply operator and using the minimum function to propagate the certainty through the ply to the conclusion avoids the problem of choosing a ply, but does not account for

the differences in these implication operators and only applies for the forward direction of the implication. However, this approach does account for the strength of the rule, and if simplicity is of paramount importance, then this approach has merit.

In the second method, the structure of the ply is intimately associated with the propagation of the certainty. This method was developed by Trillas and Valverde (reference 4), which will be referred to as the TV method. Here, a single-valued estimate of the premise validity is propagated through the rule using what is called a modus ponens generating function (MPG function). Figure 5 shows the MPG function associated with Lukasiewicz's ply. To use this function of two arguments, assign one argument as the validity of the ply and the other as the validity of the premise. The value of the function is the validity of the conclusion illustrated by the surface. This two-dimensional function is generated using the mathematical definition of the ply itself. This highly intuitive method keys on four main properties that one would like to attribute to the evidence as it propagates through a fuzzy rule (reference 4, p. 160). Using the TV notation, one first wants the propagation model to be conservative, or more precisely, to underestimate the truth of the conclusions from the available evidence. Mathematically, this means $m(v(a), v(a \rightarrow b)) \leq v(b)$, if the conclusion validity $v(b)$ was known. In practice, $v(b)$ is not known. Second, when both the premise and the consequence are absolutely certain, then the conclusion should be certain, i.e., $m(1,1) = 1$. To simplify the formulas, let $x = v(a)$ and $y = v(a \rightarrow b)$ so the last inequality can be written $m(x,y) \leq v(b)$. The third condition says that if the premise is totally uncertain, the consequence should be totally uncertain as well, or $m(0,y) = 0$. Finally, one needs a condition to guarantee that fuzzy rules can be chained, i.e., if $x \leq x'$, then $m(x,y) \leq m(x',y)$.

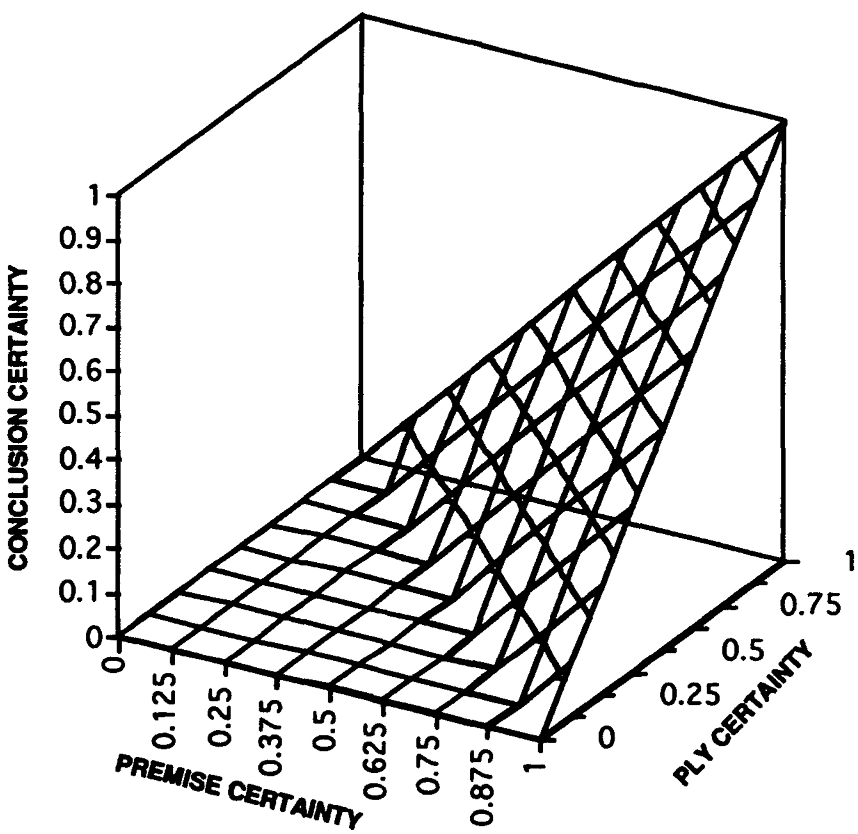


Figure 5. Modus Ponens Generating Function Associated with Lukasiewicz's Ply

In addition to these four basic properties, Hall (reference 13) has added several desirable properties that provide performance improvements. The first property $m(x,1) < 1$, $\forall x \neq 1$ says that if the premise has any uncertainty associated with it, then the conclusion cannot be absolutely certain. The second property is that the validity of the conclusion should be less than or equal to the validity of either the premise or the rule itself, i.e., $m(x,y) \leq \min(x,y)$. This property provides an upper bound to the conclusion validity. The final property lower bounds the conclusion validity away from zero, provided the premise and the ply validity are also bounded away from zero, i.e., $m(x,y) \geq 0$, if $x,y > 0$. Intuitively, this property says the validity of the conclusion should not be zero if there is some validity in the premise and the ply. This rule makes sense, but provides no sharp lower bound.

Hall (reference 13) considers several plys and their associated MPG function. One ply that satisfies six of the above seven properties is Lukasiewicz's ply defined as

$$\mu_{A(x)\to B(y)}(x,y) = \min[1, 1-\mu_A(x)+\mu_B(y)]$$

where μ_A represents the fuzzy set associated with the premise and μ_B represents the fuzzy set associated with the conclusion. The MPG function for this ply is illustrated in figure 5 and is given by $m(x,y) = \max(x+y-1,0)$. Clearly, the last property that bounds the conclusion validity away from zero is not satisfied by this MPG function.

Once the ply and its associated MPG have been determined, then the certainty propagation is simply a function evaluation. In practice, this would probably be a table look-up. Thus the structure of the implication has been included and the certainty propagation is more complex, but its implementation is straightforward and its run time is trivial.

In applying this method, the validity of compound premises such as $(a \cap c) \to b$ and $(a \cup c) \to b$ must be evaluated. Now one must find the validities $v(a \cap c)$ and $v(a \cup c)$ before one can evaluate the corresponding MPGs, e.g., $m[v(a \cup c), v((a \cup c) \to b))]$. One would like to choose $v(a \cap c) = \min(v(a), v(c))$ and it is apparent why - expediency; likewise, $v(a \cup c) = \max(v(a), v(c))$. In general, this approach cannot be justified although it is considered reasonable. Despite this lack of justification, apply this method to the braking example.

The example considered relates to an imaginary taxi driver. The fuzzy rule analyzed is the detection of a driver making a left-hand turn ahead. In the following rule, the position, slowing rate, and turn signal refer to the car ahead of the taxi:

IF the left-hand turn signal is on,
AND the car is in the left-hand lane,
AND the car has high deceleration,
THEN the car is turning left.

Figure 6 shows that the validity of each clause of the premise is determined to be 0.6, 0.8, and 0.7, respectively. Note that these values measure only the overlap of the data with the term sets. The data, a fuzzy set, represent the uncertainty of the observation. Note the left-hand signal is on with an optimistic validity 0.6; e.g., if the bright sun is in the driver's eye, it is hard to assign a higher value to this clause of the premise. This situation accounts for the width and placement of the fuzzy data. The position of the car is clearly in the left-hand lane giving an optimistic value of 0.8. Accelerations are difficult to estimate, so the third clause yields only an optimistic value of

0.7 and the fuzzy set representing the data is wider. Note that the overlap or possibility used to measure the single value of the certainty captures none of the observational uncertainty. This information will be used later in the interval-valued certainty representation.

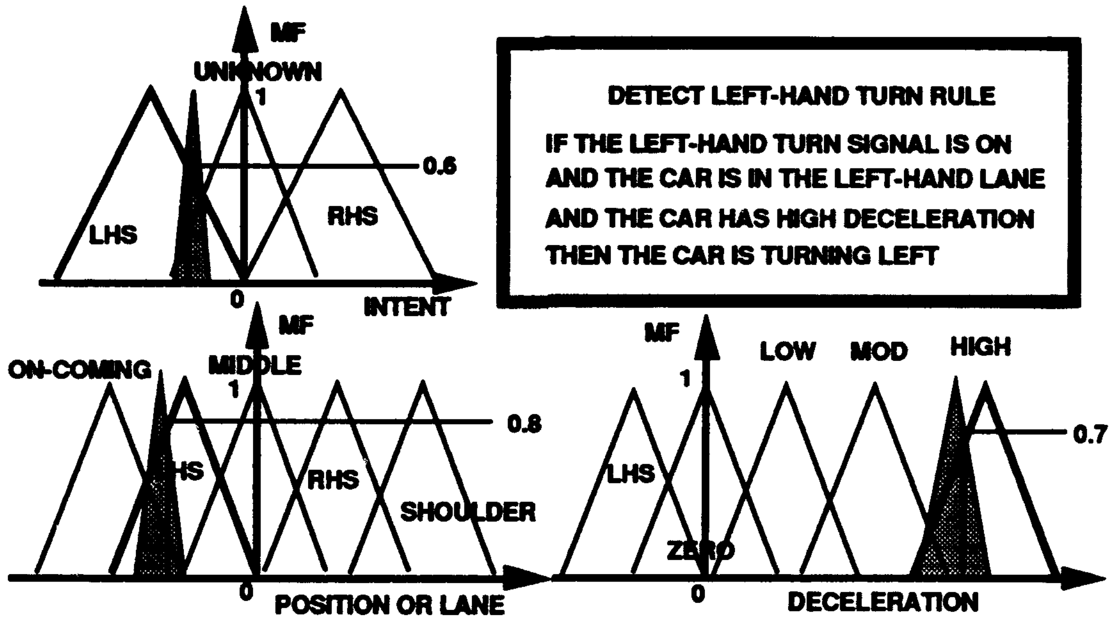


Figure 6. Detecting a Left-Hand Turn for the Car Ahead

Using the minimum of the validities as the validity of the premise yields

$$v(a) = \min(0.6, 0.8, 0.7) = 0.6.$$

The validity of the rule or ply is arbitrarily set as $v(a \rightarrow b) = 0.8$ and the ply is Lukasiewicz's ply which means $v(a \rightarrow b) = \min(1, 1 - v(a) + v(b))$ and the corresponding MPG function (reference 4) is $m(v(a), v(a \rightarrow b)) = \max(0, v(a \rightarrow b) + v(a) - 1) = 0.4$. Note that this value of the conclusion validity is even below the validity of the premise itself. This approach is not the same as simply taking the minimum of the premise and ply validity, which yields 0.6 . Moreover, the choice of the ply is important as documented by Hall (reference 13), who studied the effects of the different plys on an expert system.

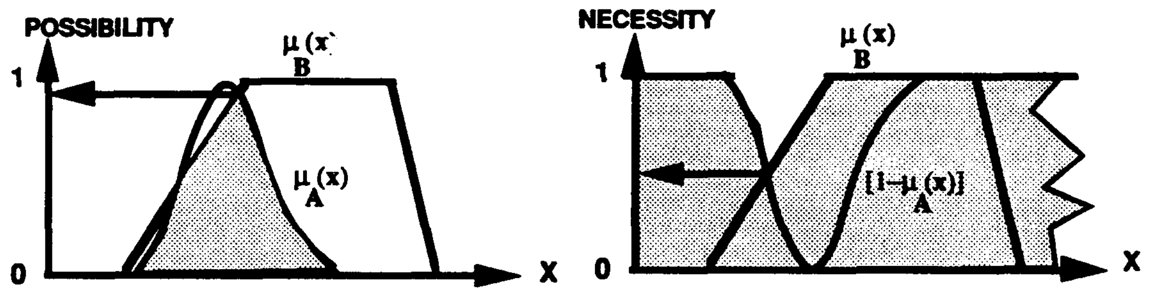
Kandel (reference 3) discusses conclusion aggregation using a method similar to an exponential learning rule. However, some of the aggregation operators studied by Klir (reference 14) might provide a more simple approach in this situation. One obvious approach is the maximum of the conclusion validities for the same conclusion. Conclusion aggregation in control is usually implemented by using the defuzzification algorithm illustrated in figure 3. When the certainty representation is no longer single-valued, conclusion aggregation becomes more of a problem.

The advantages of the two methods of evidence propagation with single-valued certainties are simplicity and practicality, which are especially important in real-time control algorithms. Disadvantages of single-valued certainties are most apparent when the data contain distributional information. Then the single-valued certainty cannot adequately take advantage of this additional information since it is not a rich enough representation. In a real-time control system, more complete certainty information is not needed because the negative feedback can quickly adjust

using only information proportional to the correct control signal. In this case, certainty information is very age dependent, meaning its utility decays rapidly with time (i.e., the correction is only relevant for the next short increment of time, until the next correction is calculated). In decision problems, the certainty of the decision has a longer time utility, so it is worthwhile to invest more resources in determining the certainty measure. The next two sections address certainty measure with higher representational complexity.

INTERVAL-VALUED CERTAINTY PROPAGATION

Interval-valued certainties associated with the premises are a natural extension of the single-valued certainties and are easily generated via the matching process: the possibility and the necessity. The possibility is an optimistic matching of data to the facts in the database, because it measures the overlap of the data and the facts and is defined as $\Pi = \sup_{x \varepsilon X} \min[\mu_A(x), \mu_B(x)]$. Also the possibility is most often used as a single-valued representation of the premise satisfaction. The necessity is a conservative certainty measure, which gives the degree of containment of data within the facts and is defined as $N = \inf_{x \varepsilon X} \max[1-\mu_A(x), \mu_B(x)]$, where A is the fuzzy data and B is the fuzzy property representing the premise (reference 15). This conservative measure provides a lower bound to the satisfaction of the property by the data. Appendix B explores other alternatives for evaluating premise satisfaction. Figure 7 illustrates the calculation of the possibility and the necessity for a test set A and a premise test set B. The interval-valued certainty discussed here is [necessity, possibility]=[N,Π].



$$POSSIBILITY = \sup_x \min[\mu_A(x), \mu_B(x)] \quad NECESSITY = \inf_x \max[1-\mu_A(x), \mu_B(x)]$$

Figure 7. Calculation of the Necessity and Possibility of A is B

The certainty of the conclusion is based on the certainty of the premise and the ply, and the certainty of the ply depends on the definition of the ply. Bonissone (reference 5-7) associates a pair of values with the ply operator that represents the strength of the implication in the forward and reverse direction. The reverse direction implication or *modus tollens* is called backward chaining by computer scientists and necessity or converse by mathematicians. (The term necessity used in this context should not be confused with the matching lower bound N called necessity; the context should make the correct meaning obvious.) The forward direction or *modus ponens* is called forward chaining by computer scientists and sufficiency by mathematicians. Thus, the forward direction strength is denoted by "suff" and the backward direction strength is denoted by "ness". These terms are used in a detachment operator along with the certainty interval of the premise to generate a certainty interval for the conclusion. A second method is to define a ply and

then propagate the certainty through the ply via the Trillas and Valverde method. Both the lower and upper bound would then be propagated separately. In the latter case, the membership function is needed for the premise and the conclusion. The second method is discussed first.

TV's notation and methodology are used to explain the bounds on the conclusion's certainty. First, note that the complement of a predicate is denoted by $n(a)$ and the validity of the complement is given by $v(n(a)) = 1 - v(a)$. *Modus ponens* furnishes the lower bound on the conclusion and *modus tollens* furnishes the upper bound on the conclusion. In the forward direction, the certainty is given by $m(v(a), v(a \to b))$, which furnishes the lower bound since $m(v(a), v(a \to b)) \leq v(b)$. The upper bound is provided by the necessity or backward part of the implication; i.e., $a \leftarrow b$ and the strength of this ply, $v(b \to a)$. Applying modus ponens in the reverse direction, one has $m[v(n(a)), v(n(a) \to n(b))] \leq v(n(b))$; and then using $v(n(b)) = 1 - v(b)$ yields an upper bound on the validity of the conclusion

$$v(b) \leq 1 - m[v(n(a)), v(n(a) \to n(b))].$$

Thus, the resultant certainty interval on the conclusion validity is $m(v(a), v(a \to b)) \leq v(b) \leq 1 - m[v(n(a)), v(n(a) \to n(b))]$. If contraposition holds, this yields $m(v(a), v(a \to b)) \leq v(b) \leq 1 - m[1 - v(a), v(b \to a)]$. If upper and lower bounds are known on the validity of the premise $v(a)$, then applying these, respectively, on the lower and upper bounds of this expression can yield a more conservative (bigger) interval.

To illustrate this method, consider the taxi driver detecting the car ahead about to make a left-hand turn. Here the premise interval-valued certainty is determined using the methods associated with the Bonissone method below, so that discussion will be deferred. For now, assume the interval-valued certainty for this compound premise is $[v_L(a), v_U(a)]$=[0.6,0.8]. The Lukasiewicz's ply is assumed with its MPG of $m(x, y) = \max(x + y - 1, 0)$, which satisfies the contraposition, and the ply certainties are assumed to be $v(a \to b)$= 0.8 and $v(b \to a)$= 0.5. In this case, the conclusion certainty interval is given by

$$[\max(v_L(a) + v(a \to b) - 1, 0), \min(1, 1 - v(b \to a) + v_U(a))],$$

which evaluates to [0.4,1.0].

To summarize, given an upper and lower bound on the premise along with the validity of the implication in both the forward and backward direction, an upper and a lower bound can be constructed on the validity of the conclusion. The construction of the conclusion validity depends on the MPG function. If only a single-valued representation of the ply is available in the forward direction, then a range for the conclusion validity can be calculated by using the MPG for the lower and upper values of the premise validity. However, it is not clear what this range means since the true validity of the conclusion may not be contained in the interval range.

Bonissone does not use the MPG used in the TV formulation; instead, the detachment operator is employed to propagate the confidence bounds through the ply. The detachment operator uses the properties of the T-norm and the S-norm, which are generalized AND and OR operators, respectively. However, the T-norm has most of the properties of the generating function, so similar conclusions hold. Recall that the forward direction $\to$ or sufficiency is denoted by "suff" and the reverse direction $\leftarrow$ or necessity is denoted by "ness". These two quantities play the role of $v(a \to b)$ and $v(b \to a)$, respectively, in the TV formulation; and the

T-norm, denoted by $T(\cdot,\cdot)$, plays the role of both the generalized fuzzy AND operator and the MPG. An example of a T-norm is the minimum function. The dual of the T-norm is the S-norm denoted by $S(\cdot,\cdot)$, which plays the role of a generalized fuzzy OR and is related to the T-norm by the equation $S(x,y) = n(T(n(x),n(y)))$. This equation is a generalized version of DeMorgan's law that is applied to the T-norm to define the S-norm with suitably defined negation operators. If $n(x) = 1 - x$, then the definition reduces to $S(x,y) = 1 - T(1-x, 1-y)$. An example of an S-norm is the maximum function. Thus, the lower bound on the confidence of the conclusion is $v_L(b) = T(suff, v_L(a))$ and the upper bound is given by

$$v_U(b) = 1 - T(v(b \to a), 1 - v_U(a)) = S(1 - ness, v_U(a))$$

where $v_L(a)$ and $v_U(a)$ are the lower and upper bounds of the premise validity, respectively. So the certainty interval of the premise propagates through the ply $[v_L(a), v_U(a)] \to [v_L(b), v_U(b)]$. Bonissone calls this operation conclusion detachment. Note two things: first, the form of the bounds derived from the T-norm is similar to the single-valued certainty propagation method with the minimum function replaced by the T-norm. However, the single-valued methods most often use the possibility or degree of overlap and not the conservative necessity of the interval-valued method. In fact, the MPG derives its conservative-ness from the fact that the implication operator must be a T-conorm. The Bonissone method is related to the TV method because they both use the T-norm in their construction; however, there is a difference in how they model the implication operation. The second fact is that Bonissone's results really apply to $v(n(a) \to n(b))$, assuming that contraposition holds.

Figure 8 illustrates how the interval-valued certainties are calculated for the three examples of single-valued certainty shown in figure 4. Here the possibility and the necessity form the upper and lower bounds on the premise certainty. Assuming again that $v(a \to b) = 1 = v(b \to a)$, the conclusion certainty intervals are illustrated as crisp sets in the figures to the right. The single-valued result, shown by the dark bar, is superimposed in the interval result to give a visual comparison. Certainty intervals can be thought of as an approximation to the the terms of the fuzzy linguistic variable TRUTH, which is discussed later in this report. The important point of this example is that the three different cases yield three distinctly different intervals of certainty. The lower bound captures the degree of containment of the data within the fuzzy premise and the upper bound models the corresponding overlap. In contrast, the single-valued certainty was simply not able to model the differences in these three cases.

For compound premises of the form

$$\bigcap_{i=1}^{n} p_i \to b,$$

the interval-valued certainty for the premise is determined by using the T-norm. If each clause p_i in the premise has a certainty interval denoted by $[a_i, A_i]$, then the premise interval is denoted by $[T(a_1, a_2, \ldots, a_n), T(A_1, A_2, \ldots, A_n)]$. Except for one special case, disjunction is handled by breaking up the implication into separate rules and then by applying conclusion aggregation to determine the certainty interval of the conclusion (aggregation is discussed in the following paragraph). Figure 9 shows how this propagation method applies with the left-turn example. Note that the interval-value is [0.4, 0.6] for the conclusion, and this interval is different than that obtained using the TV method.

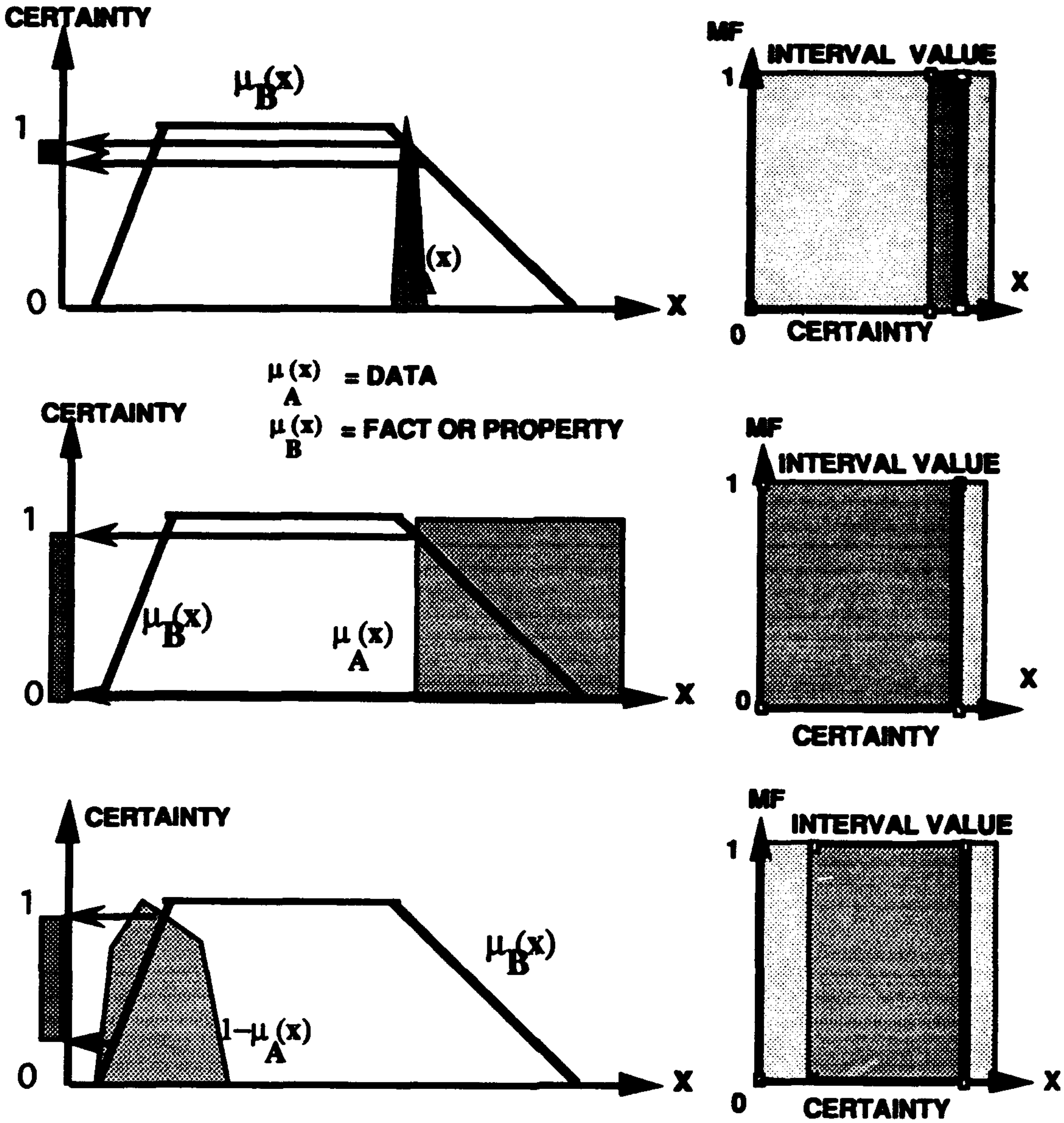


Figure 8. Examples Having the Same Single-Valued Certainty but Different Interval-Valued Certainties

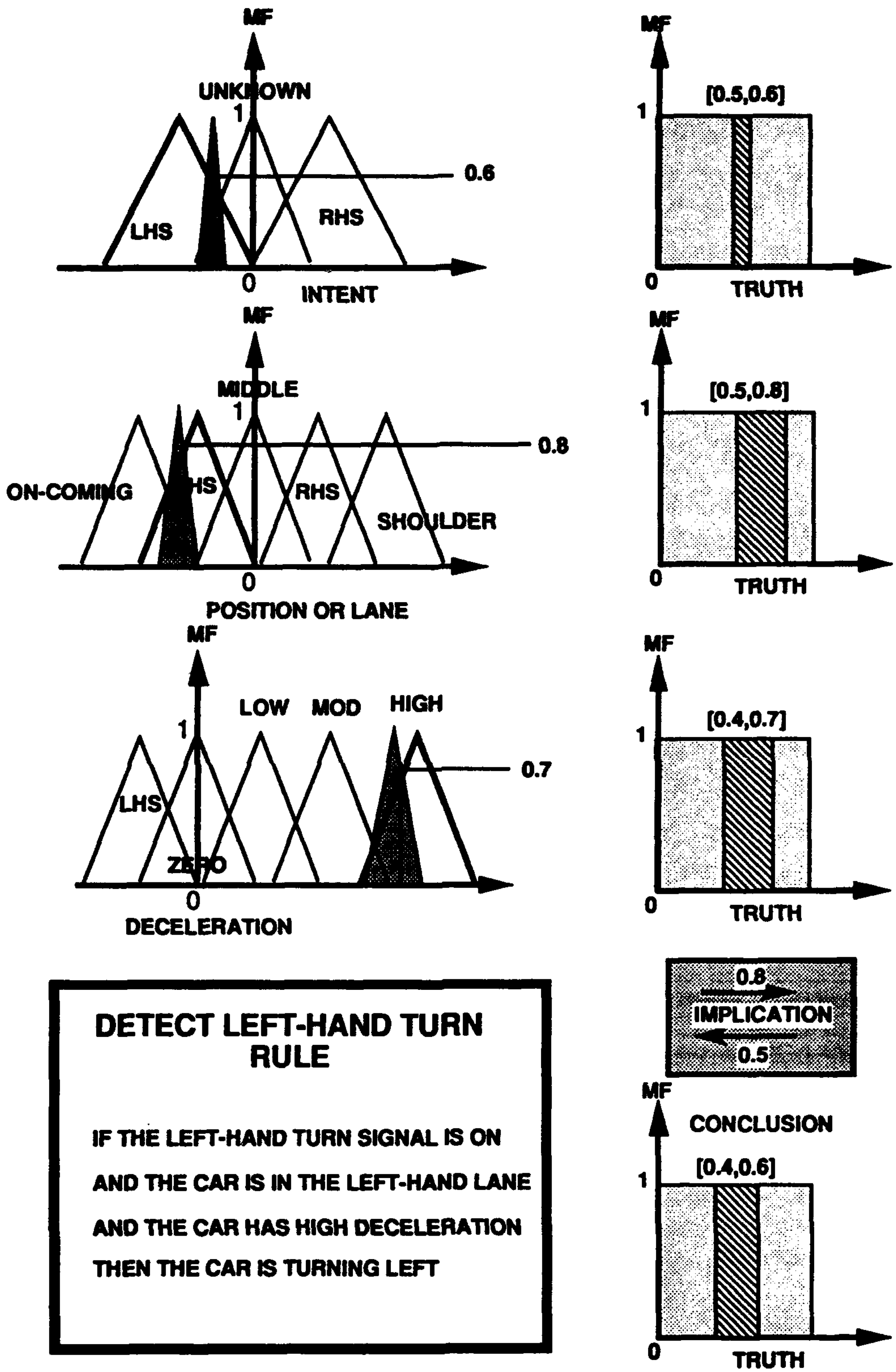


Figure 9. Detecting a Left-Hand Turn for the Car Ahead with Interval-Valued Certainty Propagation

The procedure allows the conclusion certainty to be calculated for a single rule. When several rules yield the same conclusion, each with different intervals of certainty, then the certainties must be aggregated. For multiple rules with the same conclusion C, each with corresponding certainties of $[c_i, C_i]$, the conclusion aggregation is given as

$$[S(c_1, c_2, \ldots, c_n), S(C_1, C_2, \ldots, C_n)],$$

which is a conservative method. When S is the maximum function, the aggregated lower bound is the maximum of the lower bounds and the aggregated upper bound is the maximum of the upper bounds, effectively increasing the necessity of the conclusion and widening the possibility of the conclusion. For fusion of information, a tighter bound results by applying an aggregation procedure called source consensus and yields a conclusion certainty of

$$[\max(c_1, c_2, \ldots, c_n), \min(C_1, C_2, \ldots, C_n)].$$

Source consensus reduces the spread of the certainty interval much like sampling reduces the confidence interval of an estimate.

Bonissone's method is straightforward and easily implemented, especially if the T-norm is taken as the minimum and the S-norm is the maximum. The influence of the implication in Bonissone's method is summarized with the strength of the forward and backward implications. However, the choice of the T-norm in some sense takes the place of choosing the form of the implication. Different norms are designed to model the association of the clauses within the premise (reference 5). If the clauses tend to be independent or orthogonal in nature, the product norm or $T_2(a,b) = ab$ may be appropriate. If the associations between clauses tend to be positive, then the $T_3(a,b) = \min(a,b)$ norm is appropriate. For negative associations, the norm $T_1(a,b) = \max(0, a+b-1)$ is suggested. For fuzzy rules, all the clauses may be positively associated, which means that the T_3 norm is a reasonable choice.

In the TV method of propagating evidence through the implications, the designer chooses both the implication and the T-norm before estimating the certainty interval in the premise. In Bonissone's method, the functional form of the ply does not have to be determined, but different T-norms may be used in determining the premise, the conclusion detachment, and the conclusion aggregation. The interval-valued certainty representation, more complex than the single-valued certainty, better captures the true range of certainty values from subsethood to overlap. This interval-valued representation can also be thought of as a crisp set defined as a closed interval which, in turn, may be represented as a membership function. This alternate interpretation suggests using a fuzzy set to represent the certainty, which is discussed in the next section.

FUNCTIONAL-VALUED CERTAINTY PROPAGATION

The functional-valued certainty propagation approach is a generalization of both the single-valued and interval-valued certainty representations, which uses Dubois and Prade's Fuzzy Inclusion Index (referred to as the index) to represent the satisfaction of the premise (reference 15). In binary logic, each clause of the premise must evaluate to either true or false. In fuzzy logic, clauses take on grades of truth ranging from absolutely false to absolutely true, which correspond to the false and true of binary logic. Between these two levels, many linguistic grades of truth exist, each represented as a term in the linguistic variable called TRUTH. Figure 10 indicates the values of one possible definition of the linguistic variable TRUTH (reference 16). The names true,

very true, etc., are the terms of this linguistic variable, and the collection of these terms is called the term set; the fuzzy set associated with each term or its semantic rule will also be called a term; and the index is a fuzzy set that represents the satisfaction of the premise. To propagate the certainty through the ply, one could propagate the fuzzy set pointwise using the MPG (reference 4). Another alternative is to use the Compositional Rule of Inference (CRI) and one of the many existing plys. Baldwin's investigation to this latter approach to approximate reasoning is discussed in the following paragraphs (references 17 and 18).

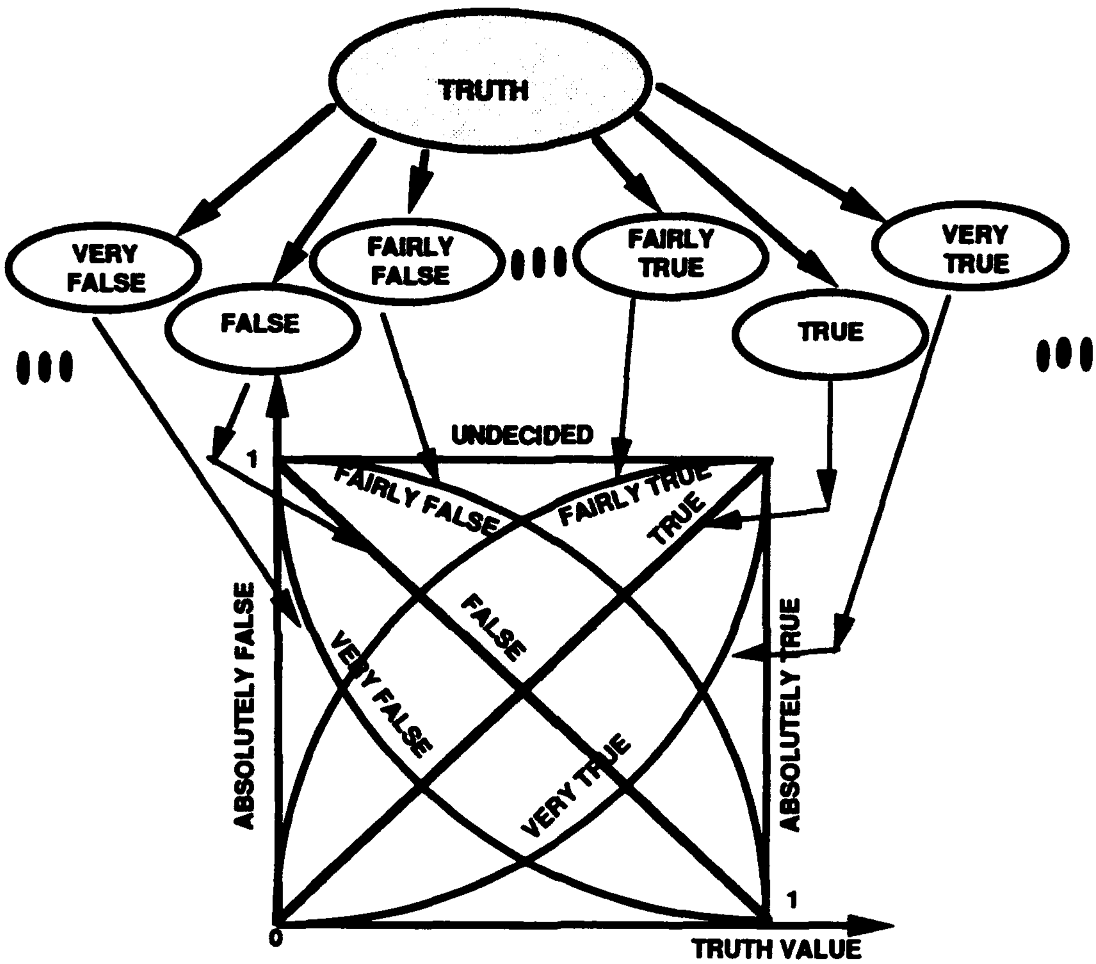


Figure 10. Linguistic Variable TRUTH

The truth of the premise is represented by a fuzzy set, not just by a single real number or by an interval of certainty. This fuzzy set can be compared with the linguistic terms of the variable TRUTH. In fact, this fuzzy set or index can be thought of as an approximation to terms like true, very true, almost true, absolutely true, false, very false, absolutely false, etc. Previous certainty measures can be thought of as approximations of the index. The single-valued certainties can be thought of as approximations to the lower or to the upper bound or to some other aspect of the index. Certainty intervals can be thought of as crisp set approximations to the Fuzzy Inclusion Index (reference 15). The Fuzzy Inclusion Index includes information from the other two representations of certainty and can be fuzzified by fitting or matching itself to the terms of the linguistic variable TRUTH, allowing a linguistic representation of certainty that may be handled as a symbolic or a numeric quantity.

Figure 11 illustrates three examples where the truth of the premise is represented as a Fuzzy Inclusion Index. A comparison of figure 11 with the interval-valued representation of figure 8 should convince the reader that the interval-valued representation is an approximation to the index. Again the single-valued certainty is represented by the dark vertical bar. The index is propagated through the Lukasiewicz implication operator using the CRI. It is known that the CRI using the Max-Min operators is an expansion operator in Turksen's terminology (reference 19) so that the level of truth in the conclusion will always be less than the premise; this will be made clearer by considering an example.

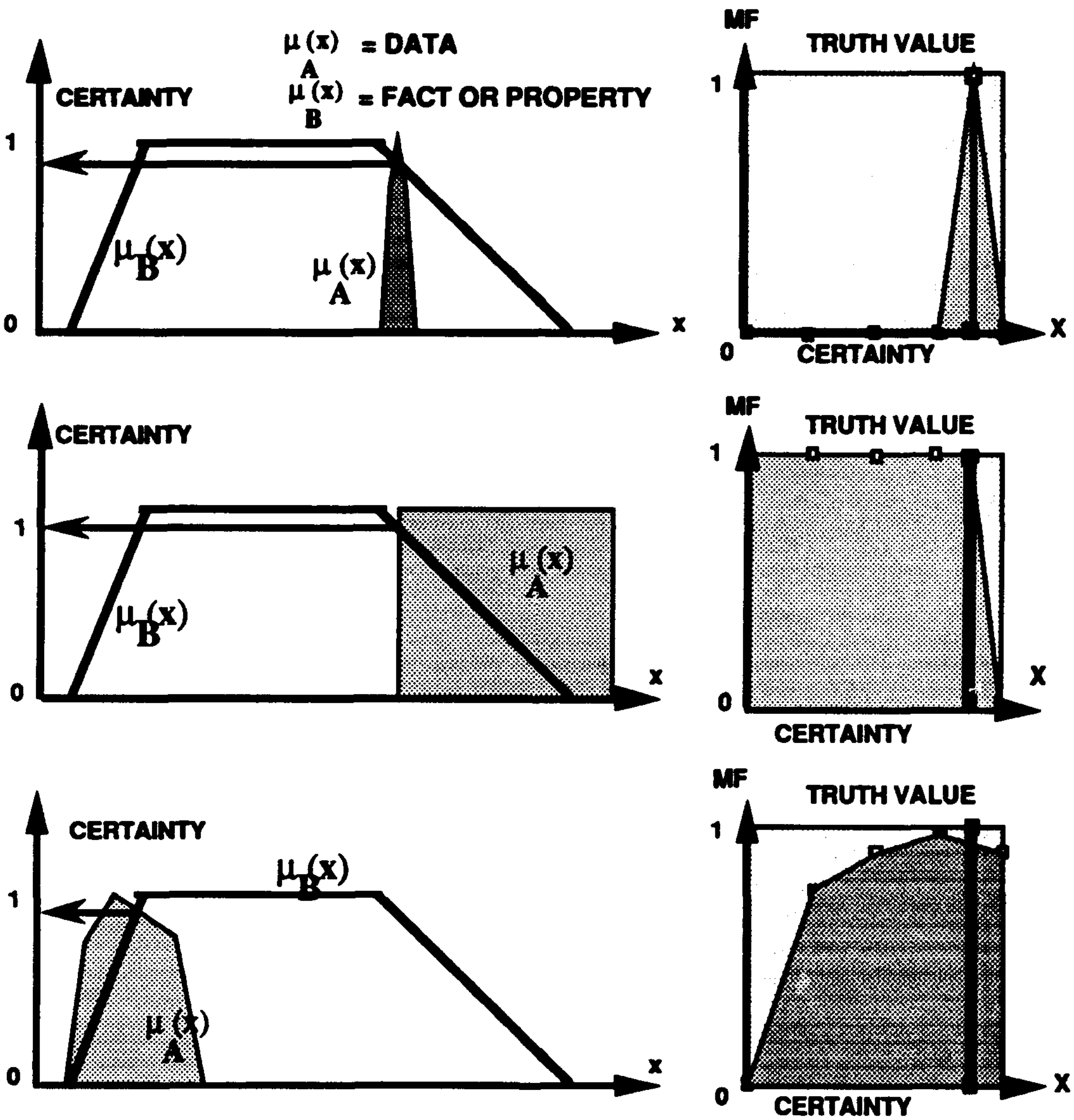


Figure 11. Examples Having the Same Single-Valued Certainty but Different Fuzzy Inclusion Indices

Consider the left-hand turn example. Figure 12 indicates how the functional-valued certainty looks for each of the premise clauses as well as the certainty for the intersection of the clauses. The certainty for the premise is calculated pairwise from the clauses in the premise. So for each clause, the index is calculated and then combined with the index of the above clause, so the second index drawn next to the clause index is the cumulative index for the premise. The functional-value certainty of the conclusion is calculated using the Lukasiewicz ply. The expansion property means that the truth of the conclusion deteriorates with each rule. Figure 12 verifies just how rapidly the truth degraded in a single passage through the ply. Baldwin (reference 17) has illustrated the effect of changing the ply on the rate of deterioration of the truth as it is propagated through the ply. In control systems, rapid truth degregation through a ply is not a problem since only one level of implication is normally needed as illustrated in the braking example. However, in multiple level implication systems such as expert systems, truth degregation through the ply is a concern that must be addressed before this method can be applied.

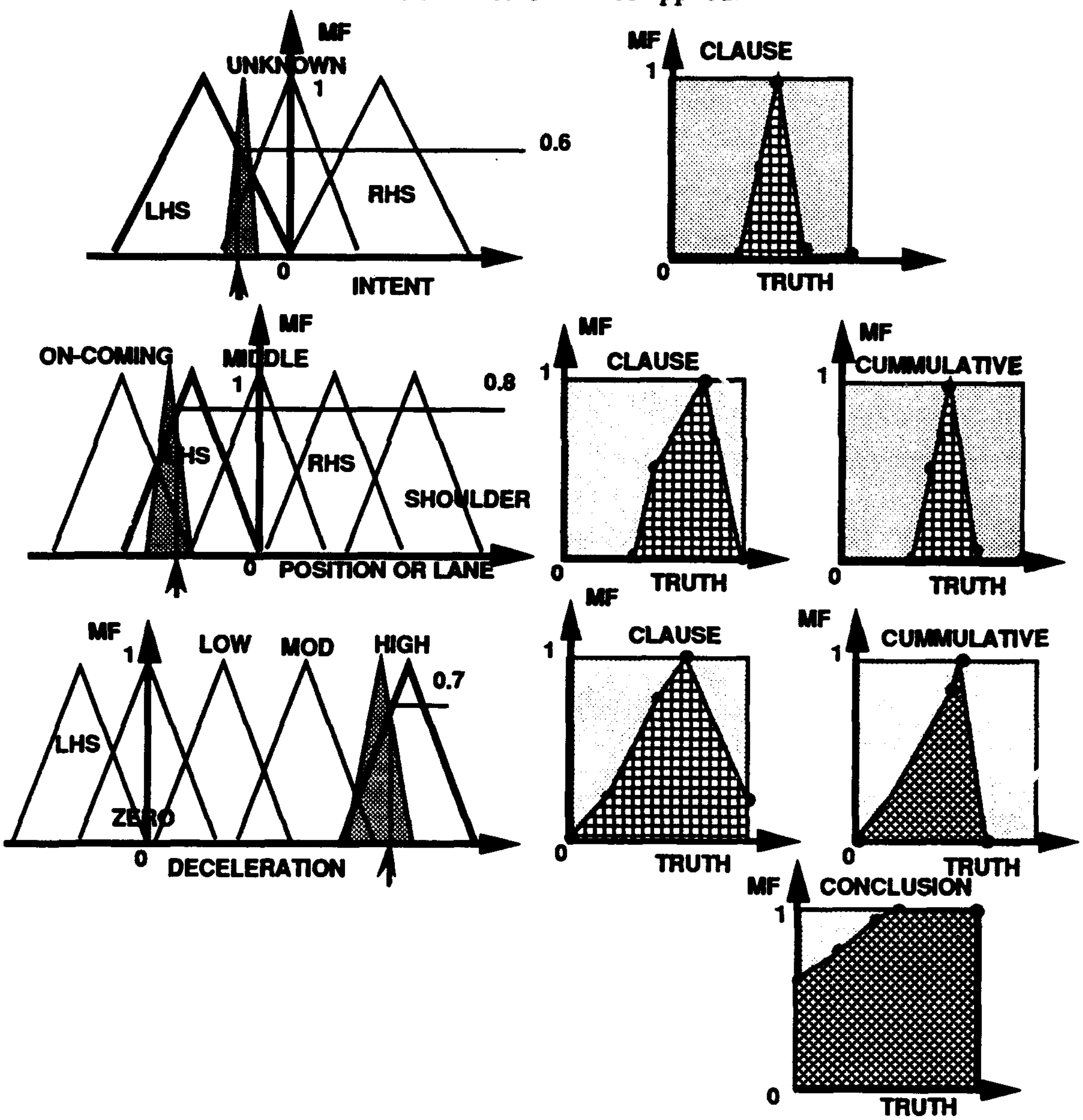


Figure 12. Detecting a Left-Hand Turn for the Car Ahead Using Truth-Functional Representation

Although the Fuzzy Inclusion Index is a much more sophisticated certainty representation, it is also an intuitively appealing notion. Propagation of this functional-valued representation through the ply markedly increases the time complexity of the algorithm. Moreover, unless the ply is designed properly, the truth deteriorates so rapidly that the method will not be effective through multiple levels of reasoning. Baldwin is well aware of both problems (reference 20). This approach needs further research to be an effective tool and to understand the tradeoff between the design of the ply and the deterioration of the truth.

SUMMARY AND CONCLUSIONS

In this report, the propagation of evidence through fuzzy rules has been studied. Evidence or data must be matched to the premise of the rule, and the strength or certainty of the match determines how strongly the conclusion is asserted. So propagating evidence is tantamount to propagating certainty. Three certainty representations have been studied along with the methods to propagate the certainty through the rule to the conclusion. The representations were ordered by increasing complexity proceeding from a single-valued representation through an interval-valued representation to a functional-valued representation. It is conceivable that all three methods could be used in a single system. Single-valued representations are limited in their ability to depict a match between the data and the rule premise; although very practical, their success probably hinges on the matching algorithm and the particular application. The interval-valued representation is better able to represent the premise certainty by capturing the spread of certainties for which the data could match the premise. Propagation through the implication is easy to compute and the aggregation of conclusion certainties is also possible. A functional-valued representation is the most general and the most difficult to implement and gives a good indication of how the data match the rule premise. Propagation through the ply is tricky and the choice of the ply-CRI method seems critical. This last method has the most promise theoretically, but also is the least practical since it is complicated and still the subject of research.

For classification problems where the feedback loop is indirect, it is recommended that a more sophisticated measure than a single-valued measure be used. In this report, only two other alternatives have been considered and the interval-valued certainty measure using Bonissone's method is the preferred choice. Propagation of evidence through fuzzy rules is still a problem of current research. No definitive solutions exist and any solution is tied to a specific application through the matching algorithm, the association of the premise clauses, and the age utility of the conclusion. Further study of this problem is clearly needed and very applicable to the problems being studied.

APPENDIX A
INTERPRETING AND APPLYING THE FUZZY INCLUSION INDEX

INTRODUCTION

The Fuzzy Inclusion Index (index) is a pattern matching measure that denotes the similarity or the degree of match of two fuzzy sets. The index itself is a fuzzy set and represents the compatibility of one fuzzy set to another fuzzy set. The set being tested will be called the test set or the data. The reference set will be called the premise or property or reference set. The Fuzzy Inclusion Index may be interpreted as

- The truth the test set possesses a property or satisfies a premise,
- The goodness of fit of the test set to the reference set,
- One of the terms of the linguistic variable TRUTH,
- The truth the test set is a subset of the reference set.

The index can be used as a general measure of the truth of the premise and can be propagated through the implication operator (ply), which will alter the shape of the index yielding a fuzzy set that represents the truth of the conclusion. However, the conclusion is then a function of the choice of the ply and the definition of the variable TRUTH. This approach is more general than single-valued or interval-valued representations of certainty.

This appendix defines the index and gives a detailed example of the calculations to construct the index. Then the interpretation of this fuzzy set as the truth that the data satisfy some premise is discussed. A parallel is drawn between the construction of the index and the statistical tests that are used if a random sample comes from a given distribution. The index is then compared with typical terms in the linguistic variable TRUTH, and a matching mechanism is described to find the "closest" term in TRUTH. The calculation of the possibility and the necessity from the index is illustrated, showing that the index contains the information that is often used as bounds to measure the compatibility of fuzzy sets. Finally, an example is given which illustrates how a fuzzy truth set propagates through an implication operator.

DEFINITIONS

The definition of the index is given in Dubois and Prade's classic text (reference 21) as an application example of the extension principle, which shows how to transform the membership function of variable x through a functional mapping. If $y = f(x)$, then the extension principle relates the membership function of y to the membership function of x. According to (reference 21) the fuzzy set induced by the function, f is

$$\mu_{\tau}(y) = \sup_{\substack{x_1,\ldots,x_r \\ x: y=f(x_1,\ldots,x_r)}} \min(\mu_{A_1}(x_1),\ldots,\mu_{A_r}(x_r))$$

provided the inverse of the point y is not empty; otherwise $\mu_{\tau}(y) = 0$, if $f^{-1}(y) = \phi$. Here the function f is $\mu_B(x)$, so the definition for one dimension becomes

$$\mu_\tau(y) = \sup_{x:y=\mu_B(x)} \mu_A(x),$$

$\forall y \varepsilon [0,1]$ provided, of course, that $\mu_B(x)$ has a nonempty inverse. Note that $\mu_\tau(y)$ is interpreted to be the compatibility of A wrt B, or A is B, or the satisfaction of the premise B by the data A. Another interpretation is as the generalization of the definition of the membership function at the point x to the membership function at the fuzzy set value B. In fact, rewriting the definition as

$$\mu_\tau(y) = \sup_{x \varepsilon \mu_B^{-1}(x)} \mu_A(x),$$

it follows that $\mu_\tau(y) = \mu_A \circ \mu_B^{-1}(y)$ where the dot is the composition of the two functions. Conceptually, this last formula is easier to interpret. For example, B=A, the $\mu_\tau(y)$ is the identity function so $\mu_\tau(y) = y$. It will be shown later that this result can be interpreted to mean that it is true that A is compatible with B, where "true" is the value taken by the linguistic variable TRUTH.

TRUTH THAT THE FUZZY TEST SET SATISFIES A PROPERTY

To illustrate the calculation and graphical construction of the Fuzzy Inclusion Index, consider two fuzzy sets: B is the reference set in the database used to represent the concept of TALL. Let A be the data representing the average height of a team. What is the truth of the statement *the team is tall?* This truth is illustrated in figure A-1 where the set B has a trapezoidal membership function and the test data have some continuous unimodal shape. The index is calculated at the five points {0, 1/4, 1/2, 3/4, 1} and the fuzzy sets are constructed from the piecewise linear approximation based on these five points. For example, for the assignment $y = 1$, the inverse of the membership function $\mu_\tau(y) = \mu_A \circ \mu_B^{-1}(y)$ has the value of $\mu_\tau(1) = \sup_{x \varepsilon [a,b]} \mu_A(x)$ as shown in figure A-1. Note the inverse of a single value for the trapezoidal form of the linguistic variable may either be a closed interval, a pair of points, or the union of two infinite intervals. The inverse of value $y = 1$ is a closed interval and the inverse of the point $y = 1/2$ gives two points {c,d}. The value is calculated from $\mu_\tau(1/2) = \max[\mu_A(c), \mu_A(d)]$ where it is obvious that $\mu_\tau(1/2) = \mu_A(c)$ since $\mu_A(d) = 0$. To calculate the index, first find the inverse of the membership value, which is a crisp set, then find the supremum of $\mu_A(x)$ over this crisp set. Thus, the more the test set A is a subset of B, the more the set possesses the same property represented by B or satisfies the premise represented by B.

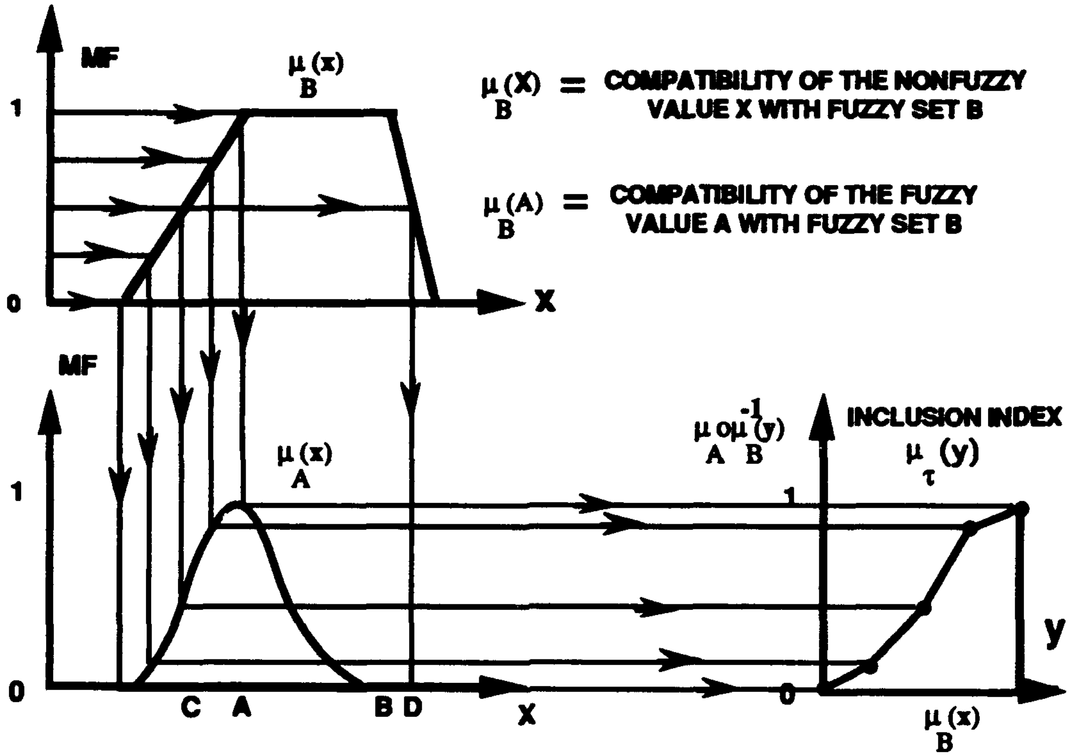


Figure A-1. Calculation of the Fuzzy Inclusion Index

The index is the truth functional that the fuzzy set A satisfies the property represented by $\mu_B(x)$ and thus the truth that A is a subset of B. To see this, compare the index with the terms of the linguistic variable TRUTH. For example, the term true is defined by Baldwin (reference 18, p. 135) as $\mu_{true}(x) = x$. The other variables are defined as powers or roots of the identity map as: $\mu_{verytrue}(x) = \mu^2_{true}(x)$, $\mu_{fairlytrue}(x) = \mu^{1/2}_{true}(x)$, and $\mu_{absolutelytrue}(x) = \delta(x-1)$ where δ is the Kronecker delta function. The definitions for false, very false, fairly false, and absolutely false follow in an obvious manner from the definition $\mu_{false}(x) = 1 - x$. Refer back to figure 10 in the text for an illustration of these definitions; note that the membership functions are only sketched, and are not plotted according to the definitions given above. The linguistic term "undecided" is $\mu_{undecided}(x) = 1, \forall x \varepsilon [0,1]$ and zero, elsewhere. Undecided means that nothing can be decided about the truth of the statement. So the index is a fuzzy set that represents the truth, which can be seen by comparing it with the terms in the linguistic variable TRUTH. However, before comparing these fuzzy sets more directly, the concept of functional distances must be, at least, defined; this is done in the following section.

GOODNESS OF FIT INTERPRETATION

Strictly speaking, fuzziness and probability measure different aspects of uncertainty. Thus any comparison between these two disciplines must be done carefully and in a way to only draw general parallels, nothing more. With this disclaimer, there are two statistical tests similar to the index: the Kolmogorov-Smirnov test and the Chi-Squared goodness of fit test. These are addressed in turn.

In this comparison, the cumulative distribution function (CDF) plays the role of the fuzzy set representing the premise. Note the comparison is already flawed since fuzzy sets can look like CDFs, but they can also look like probability density functions (PDFs) as well. In the Kolmogorov-Smirnov test, the empirical CDF is compared with a known CDF. The more identical these functions are, the more successful the test. To illustrate the parallel, consider the following example: a sequence of independent and identically distributed random variables, say $X_1,\ldots,X_n$, with CDF $F(x)$ where $F(x)$ is known. Suppose the estimate of $F(x)$, called $\hat{F}(x)$, is used. How is the goodness of fit measured? Usually, a distance measure called the Kolmogorov distance is constructed and is defined as $\sup_{x \varepsilon X}\left|\hat{F}(x)-F(x)\right|$, and this distance is tested against a threshold (reference 22). The test is rejected if the distance is too large. Other distances such as the Levy distance can be used as well. The Levy distance for two CDFs, F, and G, is defined (reference 23) to be $d_L(F,G)= \inf\{\varepsilon | \forall x, F(x-\varepsilon)-\varepsilon \le G(x) \le F(x+\varepsilon)+\varepsilon\}$. These distances will be mentioned again.

Another way to implement the test is to first define another sequence from the data samples $F(X_1),\ldots,F(X_n)$ and look at the distribution of this sequence. If it is close to the uniform CDF, then the estimate $\hat{F}(x)$ is close to $F(x)$. The uniform CDF is given by $P(X \le x) = x, \forall x \varepsilon [0,1]$, and is 0 if $x < 0$ and 1 if $x > 1$. This technique, well known in nonparametric statistics, is often used in convergence proofs, and is similar to the concept used with the index. When A and B are identical fuzzy sets, the composition of $\mu_A \circ \mu_B^{-1}(\cdot)$ becomes the identity map. When this composition is achieved, not only is A a subset of B, but also A and B are identical, meaning their membership functions are identical. Here, the sample is considered to have the same distribution as $F(x)$ if $F^{-1} \circ \hat{F}(x) \approx x$; the index measures not only subsethood but also the similarity of the two fuzzy sets or equivalently their membership function. Note the $\approx$ is due to the fact that $F(x)$ is often continuous and the empirical CDF $\hat{F}(x)$ is discontinuous by definition.

The second comparison to the index is the chi-squared goodness of fit test (reference 24) where the sampled histogram is compared with the expected histogram. This parallel is harder to draw. In the chi-squared test, one forms partitions in input space called bins and counts the number of samples that fall into each bin. The resulting plot when properly normalized gives the histogram. The same thing is done with the theoretical density function obtaining the number of expected samples in each bin and then constructing a histogram. Differences of the number of samples in each bin between the theoretical and empirical histograms are calculated, squared, properly normed, summed over the bins, and then compared with a threshold. The hypothesis that both samples came from the same distribution is rejected if the test statistic exceeds a threshold. What is being measured is the similarity of the theoretical density function $f_X(x)$ to the estimated density function $\hat{f}_X(x)$, or another way of saying this is determining the closeness of $f^{-1} \circ \hat{f}(x)$ to the identity function. Again, this comparison method is similar to the concept of the index.

COMPARING THE INDEX TO THE TERMS OF TRUTH

If the truth of a premise is to be evaluated, the index in one sense begs the question. One has a function, which is a truth functional, but no specific term or label in the term set of the linguistic variable TRUTH. Having defined the Kolmogorov distance, a more direct comparison between the index and the terms of the linguistic variable TRUTH can be made. The regularity of these TRUTH term definitions and of the index allow the application of the Kolmogorov and Levy distances to the matching process. The decisions become simplified, although the matching process has been pushed down another level, and that level is more analytically tractable. The Kolmogorov metric supplies a good comparison between the fuzzy sets, except when there are abrupt changes in the membership function because this metric does not metrize the space of CDFs. The Levy metric does metrize the CDF space and is a better choice for this matching process.

Kolmogorov and Levy metrics allow a direct comparison between the index and the terms of the linguistic variable TRUTH, so the index can be fuzzified/defuzzified to yield a quality of match. The index is a fuzzy set and the output of the matcher, but now this is to be interpreted as a linguistic term such as "very true". The metrics allow the "closest" term to be determined. To do this, compare the index with each member of the term set of TRUTH, and determine the term set with the closest to the index. If $\mu_\tau(x)$ is the inclusion index of matching the fuzzy set A to the property B and $\mu_{verytrue}(x)$ is a term of TRUTH then the distance between them is

$$d(very\ true, \tau) = \sup_{x \varepsilon [0,1]} |\mu_{verytrue}(x) - \mu_\tau(x)| .$$

This distance is constructed for each term of the linguistic variable TRUTH, and the term with the smallest distance is used as the variable value, i.e., the index has been defuzzified. Ignoring the mathematical complications, the concept of finding the closest term set is simple; choose the terms of TRUTH that look most like the index.

TRUTH TEST SET IS A SUBSET OF THE REFERENCE SET

The index measures the truth that the test set is a subset of the reference set. For example, it is known that if A is not only a subset of B ($A \subseteq B$), but also a subset of the core of B, then the index is "absolutely true." The core of B is defined as $core(B) = \{x \mid \mu_B(x) = 1\}$ and is illustrated in figure A-2a. In this illustration, the set B is a reference set and the data set is A. In figure A-2b and A-2c, the index is illustrated for data sets that are on the edge of the reference set. These examples show that the index measures the truth that A is a subset of B. When the data set A is disjoint from the reference set B, i.e., the support of B does not intersect the support of A ($\mu_A(x) > 0$ implies $\mu_B(x) = 0$), then the index is "absolutely false." Note the linguistic variable TRUTH illustrated in figure A-2 clearly lacks a complete term set. Figure A-2b shows that terms must be included with unimodal peaks near the term "absolutely false." The Beta densities that are often used as priors in Bayesian statistics would nicely augment the term "false," e.g., $\mu_{almost\ false}(x) = x^{1/2}(1-x)^5$.

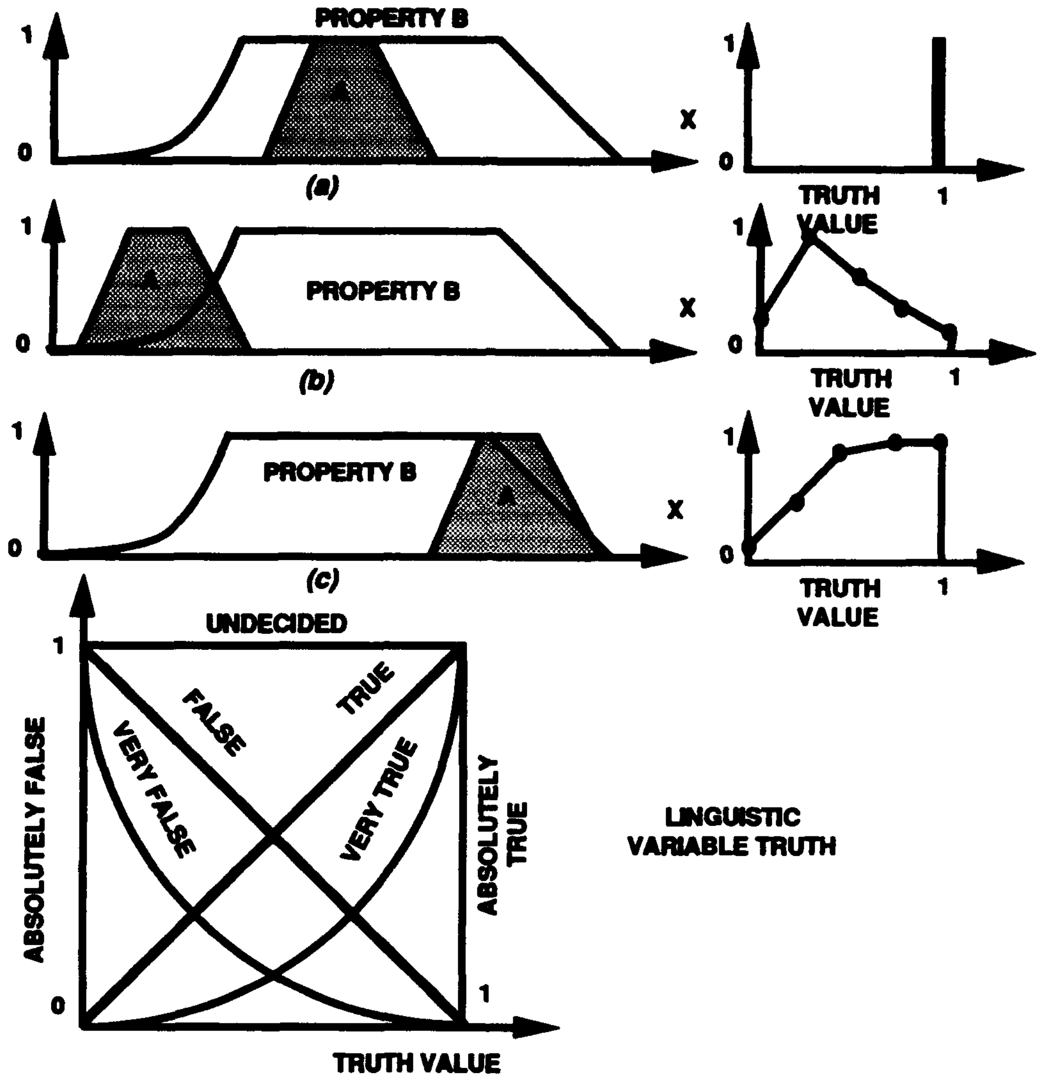


Figure A-2. Three Examples of the Inclusion Index. Property or reference set is B and the test set is A ((a) A is a subset of B, (b) A is near left edge of B, and (c) A is near right edge of B).

The calculation of the necessity and possibility from the index is found in reference 15. $\Pi(S)$ is the possibility of the statement S, where S says how well the data set A satisfies the reference set B. Recall that

$$\mu_{\tau}(y) = \sup_{x: y=\mu_B(x)} \mu_A(x).$$

The possibility A is B is given by $\Pi(S) = \sup_{v} \min[\mu_{\tau}(v), v]$ and the necessity of the statement is given by $N(S) = \inf_{v} \max[1-\mu_{\tau}(v), v]$. The graphical calculation for these fuzzy sets is illustrated in figure A-3. The possibility, which corresponds to the belief in the Dempster-Shafer

terminology, forms an upper bound to the statement being true - provided the semantic interpretation of the term true is the identity function. Likewise, the necessity, which corresponds to the plausibility in the Dempster-Shafer terminology, is a lower bound on the statement being true. So judging from figure A-3, the degrees of possibility and necessity, that S is true is given by 0.6 and 1.0. In everyday terminology, this is like saying on a scale of 0 to 10, the statement is true somewhere between a 6 and a 10.

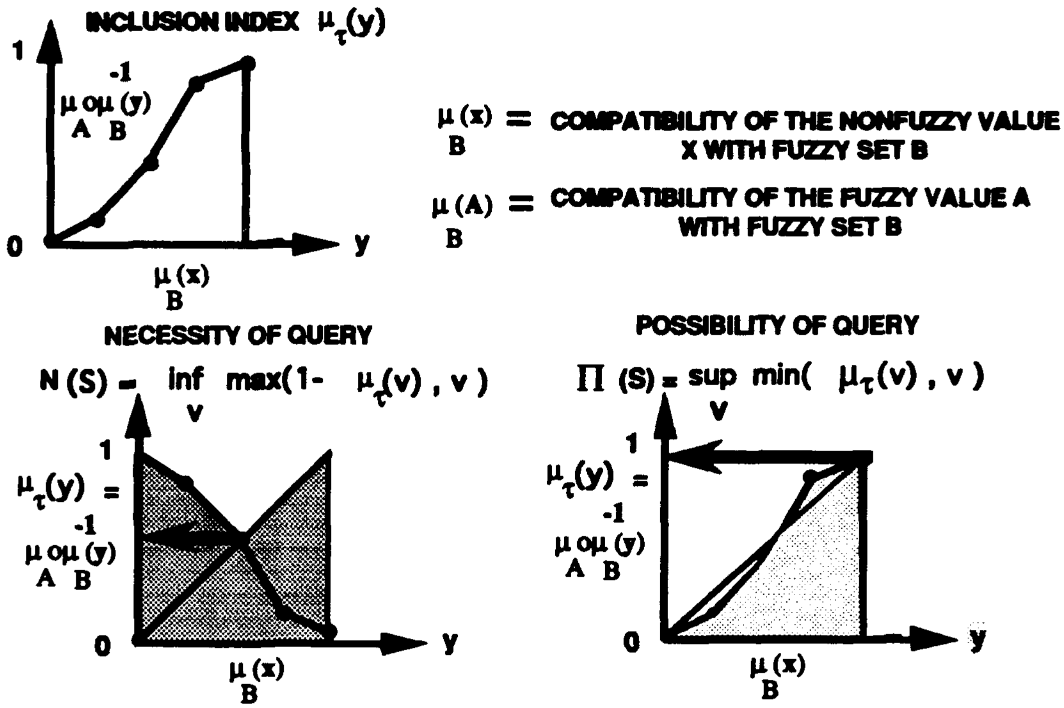


Figure A-3. Calculation of the Necessity and Possibility from the Inclusion Index

PROPAGATION OF THE INDEX THROUGH AN IMPLICATION

The index allows the truth of a premise to be represented as a fuzzy set. The CRI, developed by Zadeh, may be used to determine the truth of the conclusion, given the truth of the premise and the definition of the implication. Confusion arises because of the many ways to define the implication operation. There are a multitude of implication operators with different properties. Moreover, the way the implication is applied can differ depending on the domain of the premise and the conclusion; i.e., the ply operator may be used only on the truth values as is done in classical logic, and the truth of the conclusion is then re-interpreted on the domain where the output variable is defined. However, one may also skip the translation into truth values and work on the fuzzy data set and use the CRI from the input space directly to the output space; in fact, this is precisely what Yager does (reference 25).

To see the relationship between these two approaches, translate from Yager's approach to Baldwin's approach. The whole basis for the modus ponens is Zadeh's CRI. For modus ponens, suppose the rule is $A(x) \rightarrow B(y)$ then the CRI gives $\mu_{B'}(x) = \sup_{x \varepsilon X} \min[\mu_{A'}(x), I(x,y)]$ where $I(x,y)$ is the implication relation. For the Lukasiewicz's ply, $I(x,y) = \min(1, 1-x+y)$ or for this

case $\mu_{B'}(x) = \sup_{x \varepsilon X} \min[\mu_{A'}(x), \min(1, 1 - \mu_A(x) + \mu_B(y))]$. However, it is known from the fuzzy data A' how well the premise is satisfied or how well A' satisfies the property A. In fact, the index measures how well the property is satisfied and by definition the index is given by $\mu_{\tau_{A'}}(z) = \mu_{A'} \circ \mu_A^{-1}(z)$. Now substituting $z = \mu_A(x)$ in the CRI and observing that as the variable x ranges over its domain, z ranges over [0,1], one has

$$\mu_{B'}(y) = \sup_{z \varepsilon [0,1]} \min[\mu_{\tau_{A'}}(z), \min(1, 1 - z + \mu_B(y))].$$

Likewise, defining $w = \mu_B(y)$ and substituting this into both sides of the CRI and using the definition of the index to give $\mu_{B'} \circ \mu_B^{-1}(w) = \mu_{\tau_{B'}}(z)$ yields

$$\mu_{\tau_{B'}}(w) = \sup_{z \varepsilon [0,1]} \min[\mu_{\tau_{A'}}(z), \min(1, 1 - z + w)] \ .$$

Baldwin uses this result in his paper to relate the validity of the premise to the validity of the conclusion when both validities are represented as terms of the linguistic variable TRUTH. Figure A-4 gives a detailed example of the calculation represented by the above formula to determine the truth of the conclusion $\mu_{\tau_{B'}}(w)$ from the premise true value $\mu_{\tau_{A'}}(w)$ using Lukasiewicz's ply. Note the dashed lines in figure A-4 are for different values of w={0.0,0.25,0.5,0.75,1.0} in the above formula.

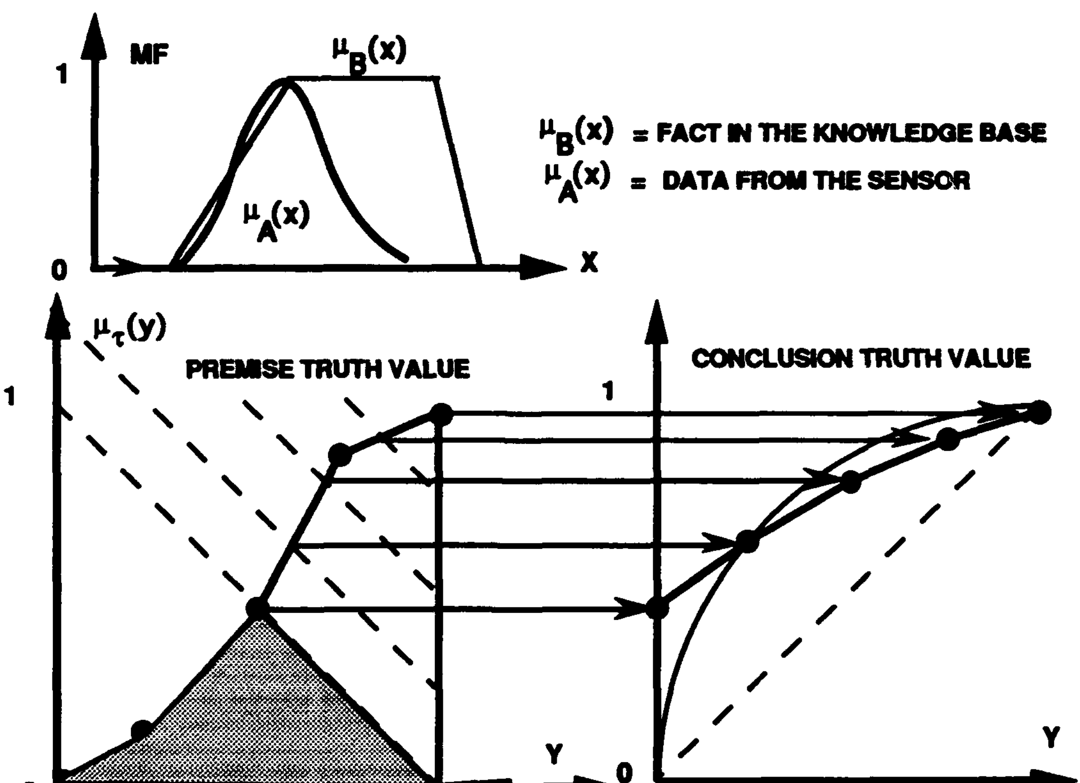


Figure A-4. Pushing the Fuzzy Inclusion Index Through the Implication

APPENDIX B
MATCHING DATA TO FACTS

Evaluating the validity of the premise and determining the validity of the implication are the subject of a vast amount of literature. In this report, the validity of the rule or implication is assumed to be known and specified by the designer of the rulebase. The validity of the premise can be calculated from the data and the clauses in the premise. In particular, the clause A is B, where A and B are both fuzzy sets, is interpreted as A has the property B. Possibility theory provides one method to calculate the validity. The clause asks how compatible the fuzzy set A is with the fuzzy set B. If $\mu_A(x)$ and $\mu_B(x)$ represent the membership functions of A and B, respectively, the one measure of the compatibility or degree of overlap of these two fuzzy sets is given by the possibility $\Pi(A \text{ is } B) = \sup_{x \varepsilon X} \min[\mu_A(x), \mu_B(x)]$. Other terms used to describe this process are pattern matching and satisfaction of the premise. No matter what the terminology, the possibility is an optimistic match of the data A to the property B. The $\Pi(A \text{ is } B)$ is interpreted as the degree that A satisfies B. If the fuzzy set A is a deterministic value $A = \{a\}$, then the possibility reduces to $\Pi(A \text{ is } B) = \mu_B(a)$, which is the degree of membership that the point a has in the fuzzy set B.

A second measure called the necessity, denoted by N, is a far more stringent measure of the concept A has the property B. In fact, it might better be interpreted to mean that A is a subset of B since it has the value 1 if and only if the set A only has support in the core of B. That is, if $\text{support}(A) = \{x \varepsilon X \mid \mu_A(x) > 0\}$ and if the $core(B) = \{x \mid \mu_B(x) = 1\}$, then necessity is 1 if and only if $\text{support}(A) \subseteq core(B)$. The necessity N is defined by the formula

$$\mathrm{N}(A \text{ is } B) = \inf_{x \varepsilon X} \max[1 - \mu_A(x), \mu_B(x)]$$

and is bounded above by the possibility $\mathrm{N}(A \text{ is } B) \leq \Pi(A \text{ is } B)$. The necessity and the possibility of the event A is B are two examples of matching algorithms. For this report, the N and Π are all that is needed to use the Bonissone results. Figure 7 showed the calculation of both the necessity and the possibility. However, these are not the only bounds that may be used.

Matching algorithms may also be based on similarity measures. Kosko's subsethood measure is one example (reference 26). In this approach, fuzzy sets are presented as vectors or points in a space I^n where I=[0,1] and n is the dimension of the fit vector or fuzzy unit vector defined as $[\mu_A(x_1), \ldots, \mu_A(x_n)]$ making up the vector. This approach works for finite fuzzy sets, which are defined as $A = \mu_A(x_1)/x_1 + \cdots + \mu_A(x_n)/x_n$. The cardinality of a fuzzy set is defined as

$$M(A) = \sum_{i=1}^{n} \mu_A(x_i)$$

where $\mu_A(x_i)$ is the membership function for the set A at the point x_i. Then the subsethood theorem of Kosko gives an expression to calculate the degree that A is a subset of B; according to Kosko (reference 26, Chapter 7), $S(A,B) = \ Degree(A \subseteq B) = \ \mu_{F(2^B)}(A \subseteq B)$ where $F(2^B)$

is the fuzzy power set of B, i.e., all the fuzzy subsets of B. With these definitions, the subsethood theorem says $S(A,B) = M(A \cup B) / M(A)$. Note that $S(\cdot,\cdot)$ denotes the subsethood measure, taking values in the interval [0,1].

Of use here is an associated concept called $SUPERSETHOOD(A,B) = 1 - S(A,B)$, which represents the concept that A is a superset of B and also the converse of the concept that A is a subset of B. Kosko calls this concept the fit violation strategy. Note $A \subset B \Leftrightarrow \mu_A(x) \leq \mu_B(x)$ so that a violation occurs when x is s.t. $\mu_A(x) > \mu_B(x)$. With this view of the problem, Kosko simply sums over all X where a violation occurs, i.e., $\mu_A(x) - \mu_B(x) > 0$. So the

$$SUPERSETHOOD(A,B) = \left[\sum_{x \varepsilon X} \max(0, \mu_A(x) - \mu_B(x)) \right] / \sum_{x \varepsilon X} \mu_A(x),$$

which is easy to calculate from the fuzzy sets. The SUPERSETHOOD(A,B) is the average number of violations of the subset property. The generalization to continuous membership functions is then clear.

SUPERSETHOOD represents the concept that A is a superset of B and also provides an efficient way of calculating the subsethood by using either summation or integration. Also of interest is the geometric interpretation of fuzzy sets as points in I^n. Reference to figure B-1 shows that the set of all fuzzy subsets or $F(2^B)$ is a closed subregion of the space I^n and thus is a compact set. Defining the subset B^* of $F(2^B)$ as the closest set to the set A, Kosko shows (reference 26, eq. 7.30 and 7.31) $d(A, F(2^B)) = \inf_{B'} \{ d(A,B') | B' \varepsilon F(2^B) \} = d(A,B^*)$. Then the subsethood can be defined as $S(A,B) = 1 - d(A,B^*) / M(A)$, which is also illustrated in figure B-1 where the distance is taken to be L^1 where $L^p = [\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|^p]^{1/p}$.

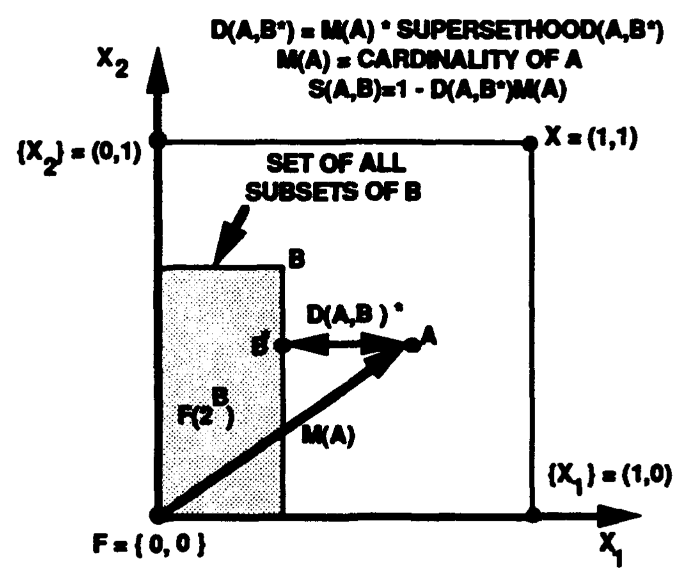


Figure B-1. Fuzzy Sets as Points and Kosko's Subsethood Measure

In effect, the S(A,B) is 1 minus the average violations of the subset property and this averaging effect makes it a good candidate for the single-valued measure of certainty.

One would think that matching a fuzzy data point to some fact or premise would just be a pattern recognition problem. When matching data to some fuzzy fact or semantic rule of the linguistic variable, the question really being asked is if this pattern or these data are a subset of the fuzzy set representing the semantic rule or the fact. In the finite case, fuzzy sets are represented as fuzzy unit vectors or fit vectors. Data fits and premise fits may not even have the same dimensions. The data must be matched to the appropriate substring of the premise fit, so one must be careful in comparing matching with pattern recognition. The data must be properly reformulated, in fact, reformulated for each pattern in the n-class problem.

Lin and Lee (references 27 and 28) have used the modified version of Kosko's subsethood to do training of the term sets for fuzzification. Here they construct a symmetric difference type of operator. In set theory, the symmetric difference $A \Delta B = (AB^C) \bigcup (A^C B)$ where A^C means the complement of A. The fuzzy counterpart of this approach is $E(A,B) = Degree(A = B) =$ $Degree(A \subseteq B \text{ and } A \supseteq B)$. The result is similar to the Entropy-subsethood theorem with $E(A,B) = M(A \bigcup B) / M(A \bigcap B)$ which is a number $E(A,B) \varepsilon [0,1]$. In terms of distances, $S(A,B) = 1 - [\, d(A,B^*) + d(B,A^*)] / M(A \bigcup B)$ When A=B, E=1, and $A \bigcap B = \phi$, then E=0. This measure is used by Lin and Lee to train terms set dynamically, and this measure is appropriate when one is testing for equality. When testing for subsethood, the previous measure S(A,B) is more appropriate. The latter quantity is suggested when trying to satisfy the premise or predicates in a premise.

REFERENCES

1. J. Bezdek, "Editorial Fuzzy Models - What are They, and Why?," IEEE *on Fuzzy Systems*, vol. 1, no. 1, February 1993, pp. 1-6.

2. R. Churchill, J. Brown, and R. Verhey, *Complex Variables and Applications*, Third Edition, McGraw-Hill, New York, pp. 181-184.

3. L. Hall and A. Kandel, *Designing Fuzzy Expert Systems*, Verlag TUV Rheinland GmbH, Cologne, Germany, 1986, pp. 1-204.

4. E. Trillas and L. Valverde, "On Mode and Implication in Approximate Reasoning," in *Approximate Reasoning in Expert Systems*, M. Gupta, A. Kandel, W. Bandler, and J. Kiszka eds., Elsevier, North-Holland, 1985, pp. 157-166.

5. P. Bonissone, "Summarizing and Propagating Uncertain Information with Triangular Norms," *International Journal of Approximate Reasoning*, vol. 1, January 1987, pp. 71-101.

6. P. Bonissone, S. Gans, and S. Decker, "RUM: A Layered Architecture for Reasoning with Uncertainty," in *Proceedings 10th International Joint Conference on Artificial Intelligence(IJCAI-87)*, Milano, Italy, 1987, pp. 891-898.

7. P. Bonissone, "Using T-norm Based Uncertainty Calculi in a Naval Situation Assessment Application," in *Third Workshop on Uncertainty in Artificial Intelligence*, Seattle, WA, July 10-12, 1987, pp. 250-261.

8. L. Zadeh, "Fuzzy Logic and Approximate Reasoning," *Synthese*, vol. 30, pp. 407-428, 1975.

9. J. Buckley, W. Siler, and D. Tucker, "A Fuzzy Expert System," *Fuzzy Sets and Systems*, vol. 20, 1986, pp. 1-16.

10. J. Buckley, "Managing Uncertainty in a Fuzzy Expert System," *International Journal of Man-Machine Studies*, vol. 29, 1988, pp. 129-148.

11. J. Buckley and D. Tucker, "Second Generation Fuzzy Expert System," *Fuzzy Sets and Systems*, vol. 31, 1989, pp. 271-284.

12. W. Siler, D. Tucker, and J. Buckley, "A Parallel Rule Firing Fuzzy Production System with Resolution of Memory Conflicts by Weak Fuzzy Monotonicity, Applied to the Classification of Multiple Objects Characterized by Multiple Uncertainty Features," *International Journal of Man-Machine Studies*, vol. 26, 1987, pp. 321-332.

13. L. Hall, "The Choice of Ply Operator in Fuzzy Intelligent Systems," *Fuzzy Sets and Systems*, vol. 34, 1990, pp. 135-144.

14. G. Klir and T. Foldger, *Fuzzy Sets, Uncertainty and Information*, Prentice Hall, Englewood Cliffs, NY, 1968.

15. D. Dubois and H. Prade, "An Introduction to Possibilistic and Fuzzy Logics," in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl, eds., Morgan Kaufmann Publishers Inc, Palo Alto, CA, 1990, pp. 743-761.

16. H. Zimmermann, *Fuzzy Set Theory and its Applications*, Second Edition, Kluwer Academic Publishers, Boston, MA, 1992.

17. J. Baldwin and N. Guild, Feasible Algorithms for Approximate Reasoning Using Fuzzy Logic, *Fuzzy Sets and Systems*, vol. 3, 1980, pp. 225-251.

18. J. Baldwin, "A New Approach to Approximate Reasoning Using a Fuzzy Logic," *Fuzzy Sets and Systems*, vol. 2, 1979, pp. 309-325.

19. Y. Tian and I. Turksen, "How to Select Combination Operators for Fuzzy Expert Systems Using CRI," *Conference Proceedings of North American Fuzzy Logic Processing Society*, (NAFIPS'92), Puerto Vallarta, Mexico, 15-17 December 1992, pp. 29-38.

20. J. Baldwin, Private conversation with J. Baldwin, Fuzzy Systems Conference, San Francisco, CA, April 1993.

21. D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, San Diego, CA, 1980.

22. J. Gibbons, *Nonparametric Statistical Inference*, McGraw-Hill, New York, 1971.

23. P. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.

24. A. Mood, F. Graybill, and D. Bose, *Introduction to the Theory of Statistics*, Third Edition, McGraw-Hill, New York, 1974.

25. R. Yager, "Approximate Reasoning as a Basis for Rule-Based Expert Systems," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-14, no. 4, July/August 1984, pp. 636-643.

26. B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, NY, 1992.

27. C. Lin and G. Lee, "Real-Time Supervised Structure/Parameter Learning for Fuzzy Neural Network," *IEEE FUZZY 92 Conference Proceedings*, 1992, pp. 1283-1291.

28. C. Lin and G. Lee, "Real-Time Supervised Structure/Parameter Learning for Fuzzy Neural Network," Engineering Research Center for Intelligent Manufacturing Systems Technical Report, Schools of Engineering, Purdue University, October 1991.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Program Executive Officer for Undersea Warfare (ASTO-B--W. Chen, ASTO-G--G. Kamilakis, ASTO-G3--LCDR Traweek)	3
Chief of Naval Operations (N872T, N872E2)	2
Chief of Naval Research (ONR-4520, ONR-4525)	2
Naval Air Warfare Center Weapons Division (Code 2158--O. McNiel, J. Hodge, A. Bergman)	3
Naval Research Laboratory (Code 5510--A. Meyrowitz)	1
Naval Surface Warfare Center Carderock Division (Code N04W--M. Stripling)	1
Commander Submarine Force Atlantic Fleet	1
Commander Submarine Force Pacific Fleet	1
Commander Submarine Development Squadron 12	1
Advanced Research Projects Agency	2
Defense Technical Information Center	12
Center for Naval Analyses	1